T.C.
İNÖNÜ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

TIBBİ GÖRÜNTÜLERİN SAÇILIM TABANLI ÖZ-DENETİMLİ ÖĞRENME İLE BÖLÜTLENMESİ

DOKTORA TEZİ

Serdar ALASU

Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı: Prof. Dr. Muhammed Fatih TALU

Ocak 2026

T.C
İNÖNÜ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

TIBBİ GÖRÜNTÜLERİN SAÇILIM TABANLI ÖZ-DENETİMLİ ÖĞRENME İLE BÖLÜTLENMESİ

DOKTORA TEZİ

Serdar ALASU
(36183619046)

Bilgisayar Mühendisliği Anabilim Dalı

Tez Danışmanı: Prof. Dr. Muhammed Fatih TALU

Ocak 2026

TEŞEKKÜR VE ÖNSÖZ

Doktora tez çalışmamın her aşamasında bilgi ve tecrübesiyle bana rehberlik eden, desteğini hiçbir zaman esirgemeyen tez danışmanım Sayın Prof. Dr. Muhammed Fatih TALU'ya,

Tez süresince görüş, öneri ve yapıcı eleştirileriyle çalışmama katkı sağlayan Doktora Tez İzleme Komitesi üyeleri Prof. Dr. Davut HANBAY ve Dr. Öğr. Üyesi Abdullah Erhan AKKAYA'ya,

Lisansüstü eğitimim süresince dostluğunu kazandığım, tez çalışmam boyunca desteği ve katkılarıyla yanımda olan Yahya ALTUNTAŞ'a

Ve en önemlisi, bu uzun ve zorlu süreçte sabrı, anlayışı ve desteğiyle her zaman yanımda olan; vazgeçmeyi düşündüğüm anlarda yeniden güç veren sevgili eşim Türkan ALASU'ya ve varlığıyla ilham kaynağı olan biricik kızım Zeynep Beren ALASU'ya,

teşekkür ederim.

ONUR SÖZÜ

Doktora tezi olarak sunduğum “Tıbbi Görüntülerin Saçılım Tabanlı Öz-Denetimli Öğrenme ile Bölütlenmesi” başlıklı bu çalışmanın bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın tarafımdan yazıldığına ve yararlandığım bütün kaynakların hem metin içinde hem de kaynakçada yöntemine uygun biçimde gösterilenlerden oluştuğunu belirtir, bunu onurumla doğrularım.

Serdar ALASU

İÇİNDEKİLER

ÇİZELGELER DİZİNİ

ŞEKİLLER DİZİNİ

SEMBOLLER VE KISALTMALAR

ESA **:** Evrişimsel Sinir Ağları (Convolutional Neural Network)

DSA **:** Dalgacık Saçılım Ağları (Wavelet Scattering Networks)

PSA **:** Parametrik Saçılım Ağları (Parametric Scattering Networks)

ÖHM **:** Öznitelik Hizalama Modülü

MLP **:** Çok Katmanlı Algılayıcı

EMA **:** Üstel Hareketli Ortalaması (Exponential Moving Avarage)

FCN **:** Tam Evrişimli Ağlar (Fully Convolutional Networks)

BYOL **:** Bootstrap Your Own Latent

ACDC **:** Automated Cardiac Diagnosis Challenge

LV **:** Sol Ventrikül (Left Ventricle)

RV **:** Sağ Ventrikül (Right Ventricle)

MYO **:** Miyakardiyum (Myocardium)

NCE **:** Noise Contrastive Estimation

ViT **:** Görsel Transformer (Vision Transformer)

K **:** Bin (x1000)

ÖZET

Doktora Tezi

TIBBİ GÖRÜNTÜLERİN SAÇILIM TABANLI ÖZ-DENETİMLİ ÖĞRENME İLE BÖLÜTLENMESİ

Serdar ALASU

İnönü Üniversitesi
Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

83+IX sayfa

2026

Danışman: Prof. Dr. Muhammed Fatih TALU

Derin öğrenme tabanlı yöntemler, son yıllarda görüntü bölütleme problemlerinde önemli başarılar elde etmiş olsa da, bu başarının temelini oluşturan denetimli öğrenme yaklaşımları yüksek miktarda piksel seviyesinde etiketlenmiş veriye ihtiyaç duymaktadır. Özellikle tıbbi görüntüleme alanında, etiketleme sürecinin uzman bilgisi gerektirmesi, zaman alıcı ve maliyetli olması, büyük ölçekli etiketli veri kümelerinin oluşturulmasını oldukça güçleştirmektedir. Bu nedenle, etiketli veriye daha az bağımlı öğrenme yaklaşımlarının geliştirilmesi, tıbbi görüntü bölütleme alanında kritik bir gereksinim haline gelmiştir.

Bu tez çalışmasında, tıbbi görüntü bölütleme problemlerinde etiketli veriye olan bağımlılığı azaltmayı amaçlayan, saçılım-tabanlı öz-denetimli öğrenme yaklaşımı önerilmiştir. Önerilen yöntemde, matematiksel olarak temellendirilmiş öteleme değişmezliği ve küçük deformasyonlara karşı kararlılık özelliklerine sahip, sabit filtrelere dayalı Dalgacık Saçılım Ağları (DSA) ile öğrenilebilir dalgacık parametresi içeren Parametrik Saçılım Ağları (PSA), karşılaştırmalı olmayan bir öz-denetimli öğrenme yöntemi olan Bootstrap Your Own Latent (BYOL) mimarisine entegre edilmiştir. Öz-denetimli ön-eğitim sonrasında elde edilen kodlayıcılar, U-Net mimarisi içerisinde denetimli olarak ince ayar yapılarak değerlendirilmiştir.

Önerilen yaklaşım, kardiyak manyetik rezonans görüntülerini içeren ACDC veri seti ile konjenital kalp hastalığına ait bilgisayarlı tomografi görüntülerinden oluşan CHD veri seti üzerinde farklı miktarlarda etiketli veri senaryoları altında gerçekleştirilmiş deneylerle test edilmiştir. Elde edilen sonuçlar, saçılım tabanlı öz-denetimli başlatma stratejilerinin, sınırlı etiketli veri koşullarında rastgele başlatma, ImageNet ön-eğitimi ve standart BYOL yaklaşımlarına kıyasla daha yüksek etiket-verimliliği sağladığını göstermektedir. Özellikle anatomik olarak zorlayıcı yapılar için PSA tabanlı yaklaşım, en yüksek bölütleme performansını sergilemiştir. Bu bulgular, saçılım dönüşümlerinin öz-denetimli öğrenme ile birlikte kullanılmasının, etiket-verimli tıbbi görüntü bölütleme için etkili bir çözüm sunduğunu ortaya koymaktadır.

Anahtar Kelimeler: Öz-denetimli öğrenme, Dalgacık saçılım ağları, Parametrik saçılım ağları, Etiket-verimli öğrenme, Tıbbi görüntü bölütleme, Kardiyak görüntü bölütleme.

ABSTRACT

Phd. Thesis

MEDICAL IMAGE SEGMENTATION WITH SCATTERING-BASED SELF-SUPERVISED LEARNING

Serdar ALASU

Inonu University
Graduate School of Nature and Applied Sciences
Department of Computer Engineering

83+IX sayfa

2026

Supervisor: Prof. Dr. Muhammed Fatih TALU

Although deep learning–based methods have achieved significant success in image segmentation problems in recent years, the supervised learning approaches underlying this success require large amounts of pixel-level annotated data. Particularly in the field of medical imaging, the need for expert knowledge in the annotation process, along with its time-consuming and costly nature, makes the curation of large-scale labeled datasets considerably challenging. Therefore, the development of learning approaches that are less dependent on labeled data has become a critical requirement in medical image segmentation.

In this thesis, a scattering-based self-supervised learning approach is proposed to reduce the dependency on labeled data in medical image segmentation problems. In the proposed method, fixed-filter Wavelet Scattering Networks (WSN), which provide mathematically grounded properties such as translation invariance and stability to small deformations, and Parametric Scattering Networks (PSN) with learnable wavelet parameters are integrated into Bootstrap Your Own Latent (BYOL), a non-contrastive self-supervised learning framework. The encoders obtained after self-supervised pretraining are evaluated through supervised fine-tuning within a U-Net architecture.

The proposed approach was evaluated through experiments conducted under different labeled data scenarios on the ACDC dataset, which consists of cardiac magnetic resonance images, and the CHD dataset, composed of computed tomography images of congenital heart disease. The obtained results demonstrate that scattering-based self-supervised initialization strategies provide higher label efficiency under limited labeled data conditions compared to random initialization, ImageNet pretraining, and the standard BYOL approach. In particular, the PSA-based approach achieves the highest segmentation performance for anatomically challenging structures. These findings indicate that the joint use of scattering transforms and self-supervised learning offers an effective solution for label-efficient medical image segmentation.

Keywords: Self-supervised learning, Wavelet scattering networks, Parametric scattering networks, Label-efficient learning, Medical image segmentation, Cardiac image segmentation.

1. GİRİŞ

1.1 Problemin Tanımı ve Motivasyon

Tıbbi görüntüleme, sağlık hizmetlerinde hastalıkların teşhisi, tedavi planlaması ve klinik süreçlerin takibi açısından hayati bir rol oynamaktadır. Manyetik rezonans görüntüleme, bilgisayarlı tomografi, ultrason ve X-ray gibi tıbbi görüntüleme teknikleri, organ ve dokulara ilişkin ayrıntılı bilgiler sunmakta ve klinik karar verme süreçlerini önemli ölçüde desteklemektedir [1]. Bu tıbbi görüntüleme teknikleriyle elde edilen veriler, yüksek boyutlu yapıları, hastalar arasında belirgin farklılıklar göstermeleri ve görüntüleme sürecinden kaynaklanan gürültüler barındırmaları nedeniyle manuel olarak analiz edilmesi zor olan veri türleridir. Bu durum, klinik iş yükünü artırmakta ve otomatik analiz yöntemlerine olan ihtiyacı daha da belirgin hale getirmektedir.

Manuel analizdeki bu kısıtlamaların üstesinden gelmek için, son yıllarda tıbbi görüntülerin otomatik analizinde derin öğrenme modelleri yaygın biçimde kullanılmaya başlanmıştır. Bu modeller, uzman bilgisine gereksinim duymadan probleme özgü öznitelikleri veriden otomatik olarak çıkarma yetenekleri sayesinde geleneksel yöntemlerden ayrışmakta ve görüntü sınıflandırması, nesne tespiti ve görüntü bölütleme gibi bilgisayarlı görü görevlerinde dikkat çekici performans göstermektedir. Bu gelişmeler, tıbbi görüntü analizinde klinik karar verme süreçlerini destekleyen otomatik sistemlerin geliştirilmesine olanak tanımıştır.

Bununla birlikte, derin öğrenme modellerinin bu başarısı büyük ölçüde denetimli öğrenme yaklaşımına dayanmaktadır. Denetimli öğrenme modelleri, genellenebilir ve ayırt edici temsiller öğrenebilmek için yüksek miktarda etiketli veriye ihtiyaç duymaktadır. Doğal görüntüler için büyük ölçekli etiketli veri kümeleri mevcutken, tıbbi görüntülerde benzer büyüklükte ve çeşitlilikte veri kümelerinin oluşturulması oldukça zordur. Tıbbi görüntülerin etiketlenmesi süreci, yüksek maliyet gerektiren, yoğun emek ve zaman isteyen bir süreçtir. Bunun temel nedeni, tıbbi verilerin etiketlenmesinde, yıllarca eğitim almış radyologlar veya patologlar gibi uzmanların bilgisine ihtiyaç duyulmasıdır. Uzmanların kısıtlı zamanı ve yüksek maliyeti, etiketleme kapasitesini sınırlamakta ve artan tıbbi görüntü hacmi ile etiketli

veri mevcudiyeti arasında giderek büyüyen bir boşluk yaratmaktadır. Bu durum, derin öğrenme modellerinin tıbbi görüntü analizinde yaygın biçimde uygulanmasının önündeki temel engellerden biri olarak öne çıkmaktadır.

Etiketli veri ihtiyacı problemi, tıbbi görüntü bölütleme görevlerinde çok daha belirgin ve kritik bir hal almaktadır [2,3]. Görüntü sınıflandırma görevlerinde görüntünün tamamı tek bir etiket ile temsil edilirken, bölütleme görevleri her bir pikselin belirli bir anatomik yapı veya patolojik bölgeye atanmasını gerektirmektedir. Bu durum, uzmanların yalnızca görüntüyü incelemesini değil, aynı zamanda tümör, organ veya lezyon sınırlarını piksel hassasiyetiyle tek tek çizmesini zorunlu kılmaktadır. Bu nedenle, piksel düzeyinde etiketleme gerektiren bu süreç, görüntü seviyesindeki etiketlemeye kıyasla belirgin ölçüde daha yüksek bir iş yükü ve maliyet oluşturmaktadır.

Özetle, derin öğrenme modellerinin görüntü bölütleme görevinde elde ettiği başarı, büyük ölçüde bu zahmetli süreçler sonucunda oluşturulan, yüksek kaliteli ve piksel seviyesinde etiketlenmiş verilerin varlığına bağlıdır [4]. Tıbbi görüntü bölütleme görevi için etiketli veri üretmenin zorluğu göz önüne alındığında, denetimli öğrenme yöntemlerinin bu görev için uygulanabilirliği önemli ölçüde kısıtlanmakta; dolayısıyla az miktarda etiketli veriyle yüksek performans gösterebilen veya etiketlenmemiş verilerden genellenebilir temsiller öğrenebilen etiket açısından verimli derin öğrenme yaklaşımlarının geliştirilmesi kritik bir gereksinim olarak ortaya çıkmaktadır.

1.2 Tezin Amacı

Bu tez çalışmasının amacı, tıbbi görüntü bölütleme problemlerinde yüksek miktarda etiketli veriye duyulan bağımlılığı azaltan, etiket açısından verimli öğrenme yaklaşımlarının geliştirilmesidir. Özellikle piksel seviyesinde etiketleme gerektiren tıbbi görüntü bölütleme görevlerinde, sınırlı etiketli veri ile etkili ve genellenebilir temsiller öğrenebilen yöntemlerin tasarlanması bu tezin temel amacını oluşturmaktadır. Bu amaç doğrultusunda, öteleme değişmezliği ve deformasyonlara karşı kararlılık gibi matematiksel avantajları olan saçılma dönüşümünün, etiketsiz veriden genelleştirilebilir temsilleri öğrenebilen öz-denetimli öğrenme metotlarına entegre eden bir ön-eğitim mimarisi önerilmiştir. Geliştirilen modelin, sınırlı etiketli veri senaryolarında, literatürdeki güncel yaklaşımlarla karşılaştırmalı performans analizlerinin yapılarak etkinliğinin kanıtlanması hedeflenmektedir.

1.3 Tezin Katkıları

Bu tez çalışması, tıbbi görüntü bölütleme alanında yaşanan etiketli veri yetersizliği problemine yönelik geliştirdiği yenilikçi saçılım dönüşümü tabanlı öz-denetimli öğrenme mimarisiyle etiket açısından verimli öğrenme yaklaşımlarına yönelik özgün katkılar sunmaktadır. Tezin temel katkıları aşağıda maddeler halinde özetlenmiştir:

- Bu çalışmada, klasik öz-denetimli öğrenme kodlayıcısının ilk aşamaları saçılım dönüşümü tabanlı katmanlar ile değiştirilerek yeni bir hibrit ön-eğitim mimarisi önerilmiştir. Bu yaklaşım sayesinde, modelin öğrenilebilir parametre sayısı azaltılmış, buna karşın derin ağların ifade kapasitesi korunmuştur.

- Saçılım dönüşümünün öteleme değişmezliği ve küçük deformasyonlara karşı kararlılık gibi matematiksel olarak temellendirilmiş özellikleri, öz-denetimli öğrenme sürecine doğrudan entegre edilmiştir. Bu entegrasyon sayesinde, tıbbi görüntü bölütleme görevi için daha kararlı ve ayırt edici bir temsil öğrenimi sağlanmıştır.

- Önerilen yaklaşım, sınırlı etiketli veri senaryolarında tıbbi görüntü bölütleme performansı açısından kapsamlı biçimde değerlendirilmiş ve farklı ön-eğitim stratejileriyle karşılaştırılmıştır. Deneysel sonuçlar, saçılım dönüşümü tabanlı öz-denetimli öğrenme yaklaşımının etiket verimliliği açısından mevcut yöntemlere kıyasla önemli avantajlar sağladığını göstermektedir.

Bu tez çalışması kapsamında aşağıda listelenen akademik yayınlar literatüre kazandırılmıştır:

- **Alasu, S., & Talu, M. F.** (2022). Mısır Tohumu Embriyolarının Bölütlenmesinde Tam Evrişimsel Ağ Tabanlı Mimarilerin Tam Bağlı Şartlı Rastgele Alanlar ile Entegrasyonu. Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi, 22(1), 100-111. https://doi.org/10.35414/akufemubid.1013047

- **Alasu, S., Talu, M. F.** (2024). Bilgisayarlı Görüde Öz-Denetimli Öğrenme Yöntemleri Üzerine Bir İnceleme. Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 12(2), 1136-1165. https://doi.org/10.29130/dubited.1201292

- **Alasu, S., Talu, M. F.** (2026). Scattering-Based Self-Supervised Learning for Label-Efficient Cardiac Image Segmentation. Electronics, 15(3), 506. https://doi.org/10.3390/electronics15030506

1.4 Tezin Organizasyonu

Bu tez çalışması, giriş bölümü dahil olmak üzere toplam beş ana bölümden oluşmaktadır. Her bölüm, araştırmanın teorik temellerini, literatürdeki yerini, geliştirilen metodolojik yaklaşımı ve elde edilen deneysel bulguları sistematik bir bütünlük içinde sunmak üzere yapılandırılmıştır. Tezin bölüm bazlı organizasyonu aşağıda özetlenmiştir:

Birinci bölümde, tez çalışmasının problemi, motivasyonu ve amacı ayrıntılı olarak açıklanmış; tıbbi görüntü bölütleme alanında etiketli veri yetersizliğinin oluşturduğu temel zorluklar ele alınmıştır. Ayrıca, tezin literatüre sunduğu katkılar özetlenmiş ve çalışmanın genel çerçevesi ortaya konulmuştur.

İkinci bölümde, tez kapsamında kullanılan yöntemlerin kuramsal arka planı sunulmaktadır. Bu bölümde, derin öğrenme tabanlı görüntü bölütleme yaklaşımları, öz-denetimli öğrenme yöntemleri ayrıntılı biçimde ele alınmaktadır. Ayrıca, Dalgacık Saçılım Ağları (DSA) ve Parametrik Saçılım Ağları (PSA) yaklaşımlarının temel prensipleri açıklanarak, önerilen yöntemin teorik temelleri oluşturulmaktadır.

Üçüncü bölümde, tıbbi görüntü bölütleme alanında öz-denetimli öğrenme yaklaşımlarının kullanıldığı ilgili çalışmalar ve saçılım tabanlı hibrit derin öğrenme modelleri sistematik bir literatür taraması çerçevesinde incelenmektedir. Bu bölümde, mevcut yöntemlerin güçlü ve sınırlı yönleri tartışılarak, tez çalışmasının literatürdeki konumu netleştirilmektedir.

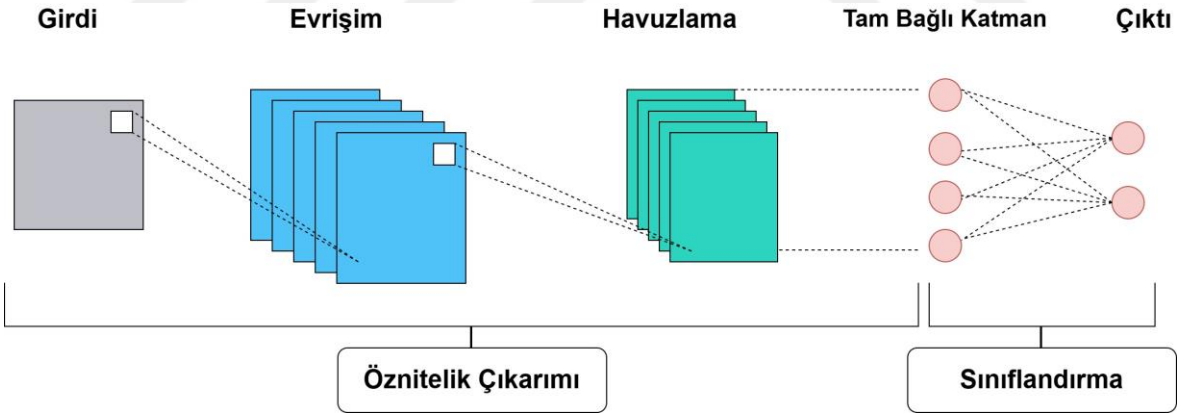
Dördüncü bölüm, tez kapsamında önerilen "Saçılım Tabanlı Öz-Denetimli Öğrenme" mimarisinin detaylı anlatımını ve bu yöntemle gerçekleştirilen deneysel çalışmaları içermektedir. Önerilen metodun tasarım süreçleri, kullanılan veri setleri, deney kurgusu ve elde edilen nicel ve nitel sonuçlar bu bölümde sunulmakta; önerilen yöntemin performansı farklı ön-eğitim stratejileriyle karşılaştırmalı olarak analiz edilmektedir.

Beşinci ve son bölümde çalışmanın literatüre sağladığı katkılar özetlenmekte ve gelecekte yapılabilecek araştırmalara yönelik önerilere yer verilmektedir.

2. KURAMSAL ARKA PLAN VE TEMEL YÖNTEMLER

2.1 Evrişimsel Sinir Ağları ve Temel Bileşenleri

Bilgisayarlı görü alanındaki klasik yaklaşımlar, ayırt edici özniteliklerin çıkarımında büyük ölçüde alan uzmanlığına dayalı olarak tasarlanan elle tanımlı (hand-crafted) özniteliklere ihtiyaç duymaktadır. Bu yaklaşımlar genellikle probleme özgü olup, farklı veri kümeleri veya görevler karşısında sınırlı bir genellenebilirlik sergilemektedir. Evrişimsel sinir ağları (ESA), bu kısıtları aşmak amacıyla geliştirilen ve veriden ayırt edici özniteliklerin otomatik olarak öğrenilmesini sağlayan derin yapay sinir ağı mimarileri olarak öne çıkmaktadır [5]. Çok katmanlı yapıları sayesinde ESA'lar, düşük seviyeli görsel örüntülerden yüksek seviyeli anlamsal temsillere uzanan hiyerarşik bir öznitelik öğrenme süreci gerçekleştirmekte ve manuel öznitelik tasarımına olan bağımlılığı büyük ölçüde ortadan kaldırmaktadır. Son yıllarda büyük ölçekli etiketli veri kümelerinin yaygınlaşması ve grafik işlem birimi (GPU) tabanlı donanımların hesaplama gücündeki önemli artış, ESA'ların etkin biçimde eğitilmesini mümkün kılarak bu mimarilerin bilgisayarlı görü alanında yaygın olarak benimsenmesini sağlamıştır [6]. Şekil 2.1'de örnek ESA mimarisinin genel yapısı gösterilmiştir.



Şekil 2.1: ESA mimarisinin genel yapısı.

ESA'ların sahip olduğu hiyerarşik öznitelik öğrenme kabiliyeti; yerel bağlantısallık (local connectivity), ağırlık paylaşımı (weight sharing) ve uzamsal alt örnekleme (pooling) olmak üzere üç temel mimari prensip üzerine inşa edilmiştir [7]. Yerel bağlantısallık, her bir nöronun giriş verisinin yalnızca sınırlı bir bölgesiyle (receptive field) etkileşime girmesini sağlayarak görüntülerdeki yerel örüntülerin etkin biçimde yakalanmasına olanak tanımaktadır. Ağırlık paylaşımı mekanizması ise aynı evrişim filtresinin görüntünün farklı uzamsal konumlarında ortak olarak kullanılmasını mümkün kılmakta; böylece hem

parametre sayısı önemli ölçüde azaltılmakta hem de aynı görsel örüntülerin konumdan bağımsız olarak tutarlı biçimde temsil edilmesi sağlanmaktadır. Bu iki prensibi tamamlayan uzamsal alt örnekleme işlemleri, genellikle havuzlama katmanları aracılığıyla gerçekleştirilmekte ve özellik haritalarının uzamsal çözünürlüğünü kademeli olarak azaltarak hesaplama yükünü düşürürken, küçük konumsal değişimlere karşı belirli bir düzeyde değişmezlik kazandırmaktadır. Bu yapısal özelliklerin bir araya gelmesi, evrişimsel sinir ağlarının yüksek boyutlu görüntü verilerini verimli, ölçeklenebilir ve genellenebilir temsillere dönüştürebilmesini sağlamaktadır.

Tipik bir ESA mimarisi, işlevsel açıdan birbirini tamamlayan iki ana bölümden meydana gelmektedir. İlk bölüm, öznitelik çıkarım bloğu olarak adlandırılan evrişim, aktivasyon ve havuzlama katmanlarından oluşmakta olup, giriş verisindeki yerel ve hiyerarşik özniteliklerin öğrenildiği aşamayı temsil etmektedir. İkinci bölüm ise düzleştirme (flattening), tam bağlı (fully connected) ve çıkış katmanlarından oluşan sınıflandırma bloğudur ve bu blokta öğrenilen öznitelikler sınıflandırma sonuçlarına dönüştürülmektedir. Her bir bileşenin işlevi ve parametre ayarları, ağın öğrenme kapasitesini, genelleme yeteneğini ve hesaplama performansını doğrudan belirlemektedir. Bu nedenle, katmanların türü, sayısı ve dizilimi hedeflenen problemin karmaşıklığına göre optimize edilmelidir. ESA mimarisini oluşturan bu temel bileşenler, takip eden alt başlıklarda ayrıntılı olarak ele alınmaktadır.

2.1.1 Evrişim katmanı

Evrişim katmanı, evrişimsel sinir ağlarının temel yapı taşı olup, giriş verisinden ayırt edici özniteliklerin çıkarılmasından sorumludur. Bu katmanda, ham görsel verideki kenar, doku ve köşe gibi yerel örüntüleri saptamak amacıyla, kare formundaki küçük ağırlık matrislerinden oluşan filtreler kullanılmaktadır. Evrişim işlemi, bu filtrelerin giriş verisi üzerinde belirli adımlarla kaydırılması ve filtrenin kapsadığı yerel bölge ile filtre ağırlıkları arasında eleman bazlı çarpım ve toplama işlemlerinin uygulanmasıyla gerçekleştirilmektedir [8]. Yerel bağlantısallık ve ağırlık paylaşımı prensiplerine dayanan bu yapı sayesinde, evrişim katmanı hem parametre verimliliği sağlamakta hem de görüntülerdeki tekrar eden örüntülerin farklı konumlarda tutarlı biçimde temsil edilmesine olanak tanımaktadır. W boyutu $k \times k$ olan bir evrişim filtresini, X giriş görüntüsünü ve b bias terimini temsil etmek üzere, evrişim işleminin matematiksel ifadesi Denklem 2.1'de verilmiştir.

$$Y(i,j) = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} W(m,n)\ X(i+m, j+n) + b \tag{2.1}$$

Evrişim katmanının giriş verisi üzerinde gerçekleştirdiği dönüşüm, yalnızca kullanılan filtrelerin öğrenilebilir ağırlıklarına değil, aynı zamanda evrişim işlemini tanımlayan birtakım yapısal parametrelere de bağlıdır. Filtre boyutu, filtre sayısı, adım uzunluğu (stride) ve doldurma (padding) gibi parametreler, elde edilen özellik haritalarının uzamsal çözünürlüğünü, ağın alıcı alanını ve hesaplama karmaşıklığını doğrudan etkilemektedir. Bu nedenle, evrişim katmanına ait parametrelerin seçimi, modelin öğrenme kapasitesi ile hesaplama verimliliği arasında önemli bir denge unsuru oluşturmaktadır.

Filtre boyutu, evrişim filtresinin giriş verisi üzerinde kapsadığı yerel alanın büyüklüğünü belirlemektedir. Genellikle 3×3 veya 5×5 gibi küçük boyutlu filtreler tercih edilmekte olup, bu yaklaşım hem parametre sayısının sınırlı tutulmasını sağlamakta hem de ağın derinleştirilmesi yoluyla daha geniş alıcı alanlara dolaylı olarak erişilmesine olanak tanımaktadır.

Filtre sayısı çıktı olarak elde edilecek öznitelik haritasının kanal sayısını belirlemektedir. Her bir filtre, giriş verisindeki belirli bir yapısal örüntüye duyarlı olacak şekilde öğrenildiğinden, filtre sayısının artırılması ağın temsil gücünü artırmakta; ancak bu durum aynı zamanda parametre sayısının da artmasına neden olmaktadır. Dolayısıyla filtre sayısının seçimi, veri kümesinin büyüklüğü ve problemin karmaşıklığı göz önünde bulundurularak yapılmalıdır.

Adım uzunluğu, evrişim filtresinin giriş üzerinde kaç piksel kaydırılarak uygulanacağını tanımlamaktadır. Daha büyük stride değerleri, öznitelik haritalarının uzamsal boyutunu azaltarak hesaplama yükünü düşürmekte ve belirli ölçüde alt örnekleme etkisi oluşturmaktadır. Ancak adım uzunluğu değerinin aşırı artırılması, ince ayrıntıların kaybolmasına ve uzamsal bilginin yetersiz temsil edilmesine yol açabilmektedir.

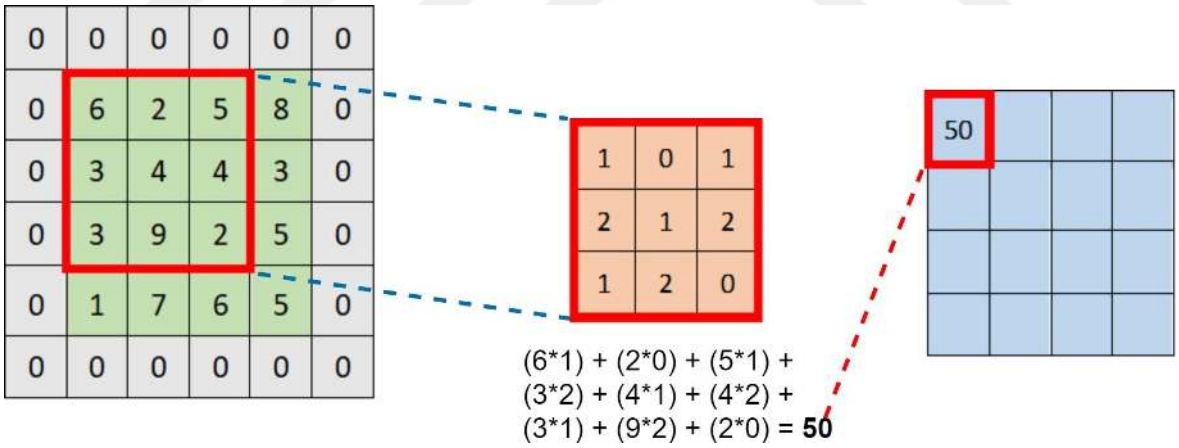
Doldurma işlemi ise evrişim sırasında giriş verisinin sınır bölgelerinde bilgi kaybını önlemek ve özellik haritalarının uzamsal boyutunu kontrol etmek amacıyla kullanılmaktadır. Doldurma işleminin uygulanmadığı durumlarda, her evrişim işlemi girişin uzamsal boyutunu küçültmekte, bu da derin ağlarda hızlı çözünürlük kaybına neden olmaktadır. Sıfır doldurma (zero-padding) gibi yaygın kullanılan doldurma yöntemleri sayesinde, giriş ve

çıkış boyutlarının korunması mümkün olmakta ve ağın derinliği artırılırken uzamsal çözünürlük daha kontrollü bir şekilde azaltılabilmektedir.

Giriş görüntüsü, filtre boyutu, doldurma ve adım uzunluğu parametreleri, evrişim işlemi sonucu elde edilen öznitelik haritasının uzamsal boyutlarını doğrudan belirleyen temel unsurlardır [8]. W giriş boyutunu, K filtre boyutunu, P doldurma miktarını ve S adım uzunluğunu göstermek üzere öznitelik haritası çıktı boyutu hesaplama formülü Denklem 2.2'de verilmiştir.

$$\text{Öznitelik Haritası Boyutu} = \left\lfloor \frac{W - K + 2P}{S} \right\rfloor + 1 \tag{2.2}$$

Burada $\lfloor \cdot \rfloor$ ifadesi, bölme işleminden sonra aşağı yuvarlamayı temsil etmektedir. Bu eşitlikler, doldurma artırıldığında çıktı boyutunun büyüyebileceğini veya korunabileceğini, adım uzunluğu artırıldığında ise çıktı boyutunun küçülerek alt örnekleme etkisi oluşacağını göstermektedir. Şekil 2.2'de iki boyutlu evrişim işleminin uygulanışı ve bunun sonucunda öznitelik haritalarının oluşturulması örneklenmiştir. Gösterimde, 4×4 boyutunda bir giriş görüntüsü üzerinde 3×3 boyutlu bir filtre kullanılmış ve adım uzunluğu 1, doldurma değeri ise 1 olarak seçilmiştir.



Şekil 2.2: Evrişim işlemi uygulamasının gösterimi.

2.1.2 Aktivasyon fonksiyonu

Evrişim katmanında gerçekleştirilen işlemler, giriş verisinin filtre ağırlıklarıyla doğrusal dönüşümlerine dayanmaktadır. Bu doğrusal yapı tek başına kullanıldığında, çok katmanlı bir ağ dahi matematiksel olarak tek bir doğrusal dönüşüme indirgenebileceğinden, karmaşık ve doğrusal olmayan ilişkilerin modellenmesi mümkün olmamaktadır [9]. Bu nedenle, evrişimsel sinir ağlarında doğrusal olmayanlık kazandırmak amacıyla evrişim

işlemini takiben aktivasyon fonksiyonları kullanılmaktadır. Aktivasyon fonksiyonları, ağın temsil gücünü artırarak giriş verisindeki karmaşık yapısal ve anlamsal örüntülerin öğrenilebilmesini sağlamaktadır.

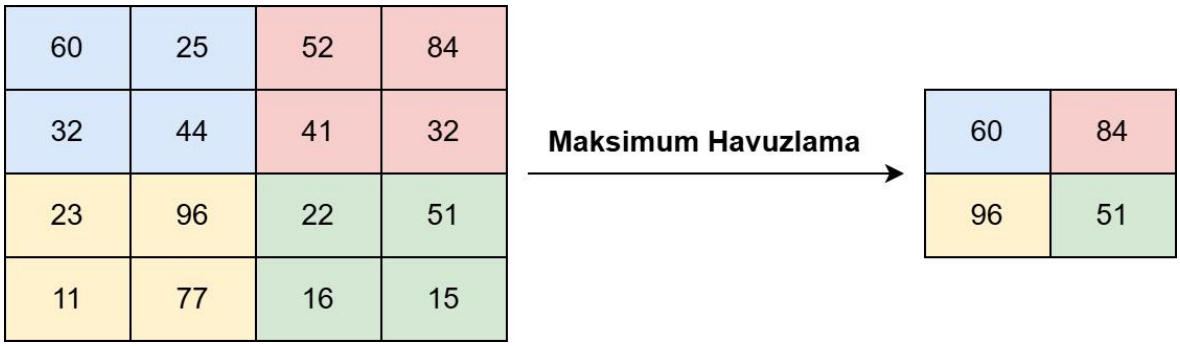
ESA mimarilerinde en yaygın kullanılan aktivasyon fonksiyonu Düzeltilmiş Doğrusal Birim (Rectified Linear Unit, ReLU) fonksiyonudur. ReLU, negatif giriş değerlerini sıfıra eşitlerken pozitif değerleri doğrudan aktarmaktadır. Bu basit doğrusal olmayan yapı, yalnızca karşılaştırma ve eşikleme işlemleri gerektirdiğinden hesaplama açısından oldukça verimlidir. Bu özellik, gradyan sönümlenmesi probleminin önemli ölçüde azaltılmasına katkı sağlamakta ve derin ağların daha kararlı ve hızlı biçimde eğitilmesine olanak tanımaktadır. Denklem 2.3'te ReLU fonksiyonu verilmiştir.

$$ReLU(x) = \max(0, x) \tag{2.3}$$

2.1.3 Havuzlama katmanı

Havuzlama, evrişimsel sinir ağlarında öznitelik haritalarının uzamsal çözünürlüğünü kademeli olarak azaltmak amacıyla kullanılan temel bir alt örnekleme (downsampling) katmanıdır. Bu katmanın temel amacı, evrişim ve aktivasyon katmanları tarafından üretilen öznitelik haritalarındaki baskın ve ayırt edici bilgileri korurken veri boyutunu küçültmek ve böylece ağın daha derin katmanlarında gerçekleştirilecek işlemlerin hesaplama yükünü azaltmaktır.

Havuzlama işlemi, genellikle belirli bir pencere boyutu ve adım uzunluğu kullanılarak, her bir pencere içerisindeki istatistiksel bir özetin hesaplanması şeklinde tanımlanmaktadır. En yaygın kullanılan havuzlama türü olan maksimum havuzlama (max pooling), her pencere içerisindeki en büyük değeri seçerek, baskın özniteliklerin korunmasını hedeflemektedir. Buna karşılık ortalama havuzlama (average pooling), pencere içerisindeki değerlerin ortalamasını alarak daha yumuşak bir temsil sunmakla birlikte bazen önemli keskin detayların kaybolmasına neden olabilmektedir. Bu iki yaklaşım arasındaki tercih, problemin doğasına ve öğrenilmek istenen temsillerin türüne bağlı olarak değişebilmektedir. Şekil 2.3'te bir maksimum havuzlama örneği gösterilmiştir. Gösterimde, 4×4 boyutunda bir giriş görüntüsü üzerinde 2×2 pencere boyutu kullanılmış ve adım uzunluğu 2 olarak seçilmiştir.

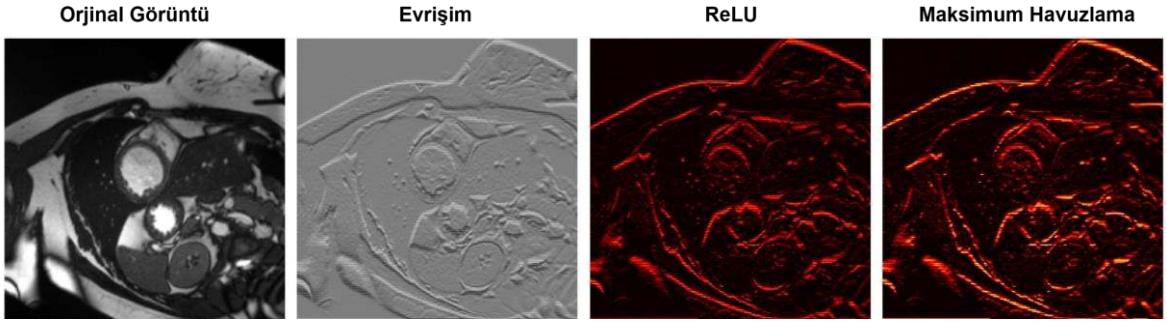


Şekil 2.3: Maksimum havuzlama işleminin gösterimi.

Havuzlama katmanlarının önemli bir etkisi, küçük konumsal değişimlere karşı belirli bir düzeyde öteleme değişmezliği (translation invariance) kazandırmasıdır [10]. Özellikle maksimum havuzlama işlemi, bir özniteliğin pencere içerisindeki tam konumundan bağımsız olarak varlığını korumasını sağlayarak, nesnelerin görüntü içerisindeki küçük yer değiştirmelerine karşı modelin daha dayanıklı hale gelmesine katkıda bulunmaktadır.

Bununla birlikte, havuzlama işlemleri uzamsal çözünürlüğü azalttığı için bilgi kaybına neden olmaktadır [11]. Özellikle piksel düzeyinde hassas konumsal bilginin kritik olduğu görüntü bölütleme gibi görevlerde, agresif havuzlama stratejileri performans kaybına yol açabilmektedir. Bu nedenle modern mimarilerde havuzlama katmanlarının sayısı ve yerleşimi dikkatle tasarlanmakta; bazı durumlarda havuzlama yerine adım büyüklüğü değerinin artırıldığı evrişim katmanları tercih edilmekte veya kaybolan uzamsal bilginin yukarı örnekleme ve atlama bağlantıları aracılığıyla telafi edilmesi amaçlanmaktadır.

Evrişim, aktivasyon ve havuzlama katmanlarının öznitelik haritaları üzerindeki etkisini daha iyi açıklayabilmek amacıyla, bu işlemlerin tek bir giriş görüntüsü üzerindeki ardışık çıktıları Şekil 2.4'te görselleştirilmiştir.



Şekil 2.4: Evrişim, aktivasyon fonksiyonu ve maksimum havuzlama işlemlerinin çıktılarının görselleştirilmesi.

2.1.4 Düzleştirme katmanı

Evrişim, aktivasyon ve havuzlama katmanlarından oluşan öznitelik çıkarım bloğunun temel çıktısı, giriş görüntüsündeki uzamsal hiyerarşiyi temsil eden Yükseklik × Genişlik × Kanal Sayısı boyutlu öznitelik haritalarıdır. Ancak, sınıflandırma aşamasında kullanılan tam bağlantılı katmanlar, yapıları gereği tek boyutlu giriş vektörleri ile çalışmaktadır. Bu nedenle, evrişim tabanlı öznitelik çıkarım aşaması ile tam bağlı katmanlar arasında, çok boyutlu öznitelik haritalarını tek bir sütun vektörüne dönüştüren düzleştirme (flatten) işlemi uygulanmaktadır [12].

Düzleştirme işleminde, her bir öznitelik haritasındaki hücre değerleri sırasıyla okunarak uç uca eklenir ve sonuçta tüm yerel bilgileri içeren yüksek boyutlu tek boyutlu bir öznitelik vektörü elde edilir. Bu işlem sırasında herhangi bir öğrenilebilir parametre kullanılmamakta, yalnızca verinin yeniden şekillendirilmesi gerçekleştirilmektedir. Bu işlem sonucunda elde edilen vektörün boyutu, öznitelik haritasının yüksekliği, genişliği ve kanal sayısının çarpımıyla belirlenmektedir. Bu vektör, ağın sonraki aşamalarında yer alan tam bağlı katmanlar tarafından işlenerek nihai tahmin çıktısına dönüştürülmektedir.

Düzleştirme işlemi, evrişimsel temsillerin global düzeyde birleştirilmesine olanak tanırken, aynı zamanda uzamsal yapı bilgisinin korunmamasına yol açmaktadır. Özellikle piksel düzeyinde konumsal bilginin kritik olduğu görüntü bölütleme gibi görevlerde, bu durum önemli bir sınırlılık olarak ortaya çıkmaktadır.

2.1.5 Tam bağlı katman

Tam bağlı katman, genellikle ESA mimarilerinin son aşamasında yer alan ve öznitelik çıkarım aşamasında öğrenilen temsillerin nihai tahmin çıktısına dönüştürüldüğü bölümdür. Düzleştirme işlemi sonrasında elde edilen tek boyutlu öznitelik vektörü, tam bağlı katmanlara giriş olarak verilmekte ve bu katmanlar aracılığıyla sınıflandırma veya regresyon gibi görevler gerçekleştirilmektedir. Tam bağlı olarak adlandırılmasının temel nedeni, bu katmandaki her bir nöronun bir önceki katmandaki tüm aktivasyon birimleri ile bağlantılı olmasıdır. Matematiksel olarak, $\mathrm{W} \in R^{M \times N}$ ağırlık matrisi, $\mathrm{b} \in R^{M}$ bias vektörünü ve y katman çıktısını temsil etmek üzere, bir tam bağlı katman, giriş vektörü $\mathrm{x} \in R^{N}$ için Denklem 2.4'te verilen doğrusal dönüşümü gerçekleştirmektedir.

$$\mathrm{y} = \mathrm{Wx} + \mathrm{b} \tag{2.4}$$

Bu katmanlar, evrişim katmanlarının yakaladığı yerel ve karmaşık örüntüler arasındaki global ilişkileri modelleyebilmektedir. Bu sayede, giriş görüntüsünün genel yapısına ilişkin yüksek seviyeli anlamsal çıkarımlar yapılabilmektedir. Ancak bu yapı, parametre sayısının hızla artmasına neden olmakta ve özellikle yüksek boyutlu öznitelik vektörleri söz konusu olduğunda hesaplama maliyetini ve aşırı öğrenme riskini artırmaktadır. Bu nedenle, tam bağlı katmanlar genellikle ağın son aşamalarında sınırlı sayıda kullanılmakta veya dropout teknikleri [13] ile desteklenmektedir.

Bununla birlikte, tam bağlı katmanların kullanımı uzamsal yapı bilgisinin açık biçimde korunmaması gibi önemli bir sınırlılığı da beraberinde getirmektedir. Düzleştirme işlemiyle birlikte, öznitelik haritalarındaki konumsal ilişkiler kaybolmakta ve piksel düzeyinde hassasiyet gerektiren görevlerde performans düşüşü gözlenebilmektedir. Bu nedenle, klasik tam bağlı katmanlara dayalı ESA mimarileri, görüntü sınıflandırma gibi global tahmin gerektiren görevlerde başarılı sonuçlar üretirken, anlamsal bölütleme gibi yoğun tahmin gerektiren görevler için uygun değildir. Modern kodlayıcı–kod çözücü tabanlı mimarilerde düzleştirme ve tam bağlı katman çoğunlukla kullanılmamakta, bunun yerine uzamsal yapıyı koruyan tamamen evrişimsel (fully convolutional) tasarımlar tercih edilmektedir.

2.1.6 Çıkış katmanı

Çıkış katmanı, tam bağlı katmanlar tarafından üretilen ve logit olarak adlandırılan ham çıktıları, problem tanımına uygun bir çıktı biçimine dönüştürmektedir. Bu katman, ağın son katmanı olup, öğrenilen temsillerin sınıflandırma veya regresyon gibi görevler için yorumlanabilir nicel değerlere dönüştürüldüğü aşamayı temsil etmektedir. Bu katmanda kullanılan aktivasyon fonksiyonu, doğrudan ele alınan problemin türüne bağlı olarak belirlenmektedir.

Çok sınıflı sınıflandırma problemlerinde, çıkış katmanı sınıf sayısı kadar nörondan oluşmakta ve her bir nöron ilgili sınıfa ait logit değerini üretmektedir. Bu logit değerleri, softmax aktivasyon fonksiyonu aracılığıyla normalize edilerek her bir sınıfa ait olasılık dağılımı elde edilmektedir [14]. Çıkış katmanındaki i. nöronun ürettiği logit değeri z_i ile gösterilmek üzere, i. sınıf için tahmin edilen olasılık değerinin (P_i) hesabı Denklem 2.5'te verilmiştir.

$$p_i = \frac{e^{z_i}}{\sum_j e^{z_j}} \tag{2.5}$$

Bu işlem sonucunda, çıktı vektöründeki her bir bileşen [0,1] aralığında bir olasılık değeri almakta ve bu değerlerin toplamı bire eşit olmaktadır. Nihai sınıf tahmini, en yüksek olasılığa sahip sınıfın seçilmesiyle gerçekleştirilmektedir.

İkili sınıflandırma problemlerinde ise çıkış katmanı tek bir nörondan oluşmakta ve sigmoid aktivasyon fonksiyonu kullanılmaktadır. Sigmoid fonksiyonu, çıkış değerini olasılıksal bir yoruma uygun şekilde [0,1] aralığına sıkıştırarak, ilgili örneğin pozitif sınıfa ait olma olasılığını ifade etmektedir. Regresyon problemlerinde ise çıktı katmanında doğrusal bir aktivasyon fonksiyonu tercih edilmekte ve modelin sürekli değerler üretmesi sağlanmaktadır.

2.2 Evrişimsel Sinir Ağları Mimarileri

2.2.1 AlexNet

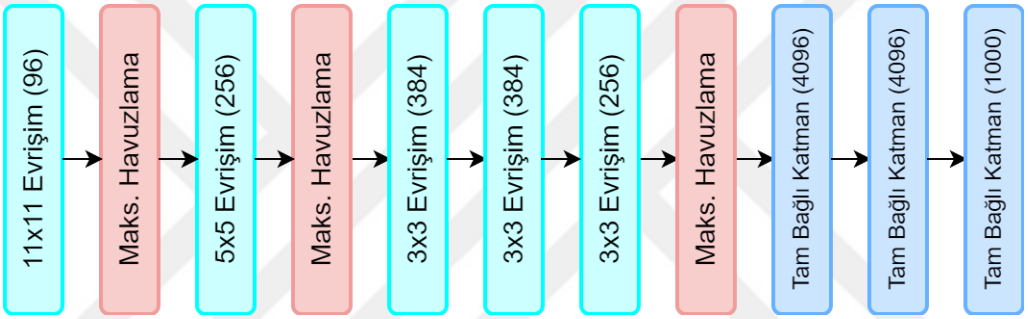
AlexNet [15], derin evrişimsel sinir ağlarının büyük ölçekli görsel tanıma problemlerinde etkin bir şekilde kullanılabileceğini ilk kez ortaya koyan öncü mimarilerden biridir. 2012 yılında ImageNet Görsel Tanıma Yarışması'nda (ILSVRC) elde ettiği belirgin başarı ile birlikte, derin öğrenme yaklaşımlarının bilgisayarla görü alanında yaygınlaşmasında kritik bir dönüm noktası olmuştur. Bu çalışma, klasik makine öğrenmesi tabanlı yaklaşımların yerini, uçtan uca öğrenilebilen derin mimarilere bırakmasının önünü açmıştır.

AlexNet mimarisi, beş evrişim ve üç tam bağlantılı katmandan oluşan derin bir yapı sunmaktadır. AlexNet'in başarısı sadece katman sayısındaki artışa değil, aynı zamanda eğitim sürecini getirdiği yenilikçi yaklaşımlara dayanmaktadır. Bu yaklaşımların başında, o döneme kadar yaygın olarak kullanılan sigmoid ve hiperbolik tanjant fonksiyonları yerine ReLU aktivasyon fonksiyonunun tercih edilmesi gelmektedir. ReLU kullanımı, gradyan kaybı problemini azaltarak ağın çok daha hızlı yakınsamasını sağlamış ve derin ağların eğitilebilirliğini önemli ölçüde artırmıştır.

Modelin yaklaşık 60 milyon parametre içeren devasa yapısı, beraberinde aşırı öğrenme (overfitting) ve yüksek hesaplama maliyeti gibi iki temel zorluğu getirmiştir. Bu problemleri aşmak amacıyla, çalışmada eğitim sırasında rastgele seçilen nöronların devre dışı bırakıldığı dropout tekniği ve veri setini sentetik olarak zenginleştiren veri artırma yöntemleri önerilmiştir. Ayrıca, dönemin donanımsal yetersizliklerini aşmak amacıyla model, iki ayrı GPU üzerinde paralel olarak eğitilecek şekilde tasarlanmıştır. Bu donanımsal

paralelleştirme stratejisi, modern derin öğrenme kütüphanelerinin ve çoklu GPU kullanımının temelini oluşturmuştur. İlk evrişim katmanında kullanılan 11×11 gibi geniş filtre boyutları, görüntüdeki kaba öznitelikleri yakalarken, takip eden katmanlardaki daha küçük filtreler nesne detaylarının hiyerarşik bir biçimde öğrenilmesine imkan tanımıştır.

Sonuç olarak AlexNet, ReLU aktivasyon fonksiyonu, dropout stratejisi ve GPU destekli eğitim gibi günümüzde standart haline gelmiş birçok bileşeni bünyesinde barındırarak, geleneksel öznitelik mühendisliğine dayalı yaklaşımların yerini veriden uçtan uca öznitelik öğrenebilen derin öğrenme modellerinin almasına öncülük eden bir mimari olmuştur. Bu yönüyle AlexNet, daha derin ve yapısal olarak daha gelişmiş mimarilerin geliştirilmesine zemin hazırlayan öncü bir çalışma olarak değerlendirilmektedir. Şekil 2.5'te AlexNet mimarisinin genel yapısı gösterilmiştir.



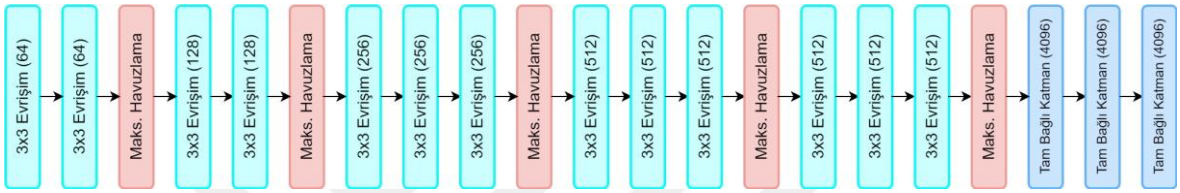
Şekil 2.5: AlexNet mimarisinin genel yapısı.

2.2.2 VGG

VGG [16] mimarisi, AlexNet ile ortaya konan derin evrişimsel sinir ağı yaklaşımını daha sade ve sistematik bir mimari tasarım anlayışıyla ileriye taşıyan önemli bir çalışmadır. Oxford Üniversitesi Görsel Geometri Grubu (Visual Geometry Group) tarafından önerilen bu mimari, AlexNet'in karmaşık ve farklı boyutlardaki filtre yapısının aksine, ağın tüm katmanlarında sabit ve küçük ölçekli filtreler kullanarak mimari tasarımda bir standardizasyon getirmiştir. VGG, derinliğin ağ performansı üzerindeki etkisini inceleyen ilk sistematik çalışmalardan biri olarak literatürde önemli bir yere sahiptir.

VGG mimarisinin temel tasarım prensibi, büyük boyutlara sahip evrişim filtreleri yerine, ardışık olarak kullanılan 3×3 boyutundaki evrişim filtreleri ile daha derin ağların inşa edilmesidir. Bu yaklaşım sayesinde, daha az parametre ile daha geniş bir alıcı alan (receptive field) elde edilmekte ve doğrusal olmayanlık sayısı artırılarak daha zengin hiyerarşik temsiller öğrenilebilmektedir.

Literatürde en yaygın kullanılan VGG varyantları olan VGG-16 ve VGG-19, sırasıyla 16 ve 19 adet öğrenilebilir katmandan oluşmakta olup, evrişimsel katmanların bloklar halinde gruplanması ve her blok sonunda maksimum havuzlama işlemi uygulanması esasına dayanmaktadır. Mimari, evrişimsel katmanlardan sonra tam bağlantılı katmanlar ile sonlandırılmakta ve sınıflandırma çıktısı üretilmektedir. Bu düzenli ve tekrarlanabilir blok yapısı, VGG'yi mimari analiz ve karşılaştırmalar için ideal bir referans modeli haline getirmiştir. Bununla birlikte, VGG mimarisinin en önemli sınırlamalarından biri, yüksek parametre sayısı ve buna bağlı bellek ve hesaplama maliyetidir. Özellikle tam bağlantılı katmanlar, modeldeki parametrelerin önemli bir bölümünü oluşturmakta ve bu durum eğitim sürecinin zorlaştırmaktadır. VGG-16 mimarisinin genel yapısı Şekil 2.6'da gösterilmiştir.

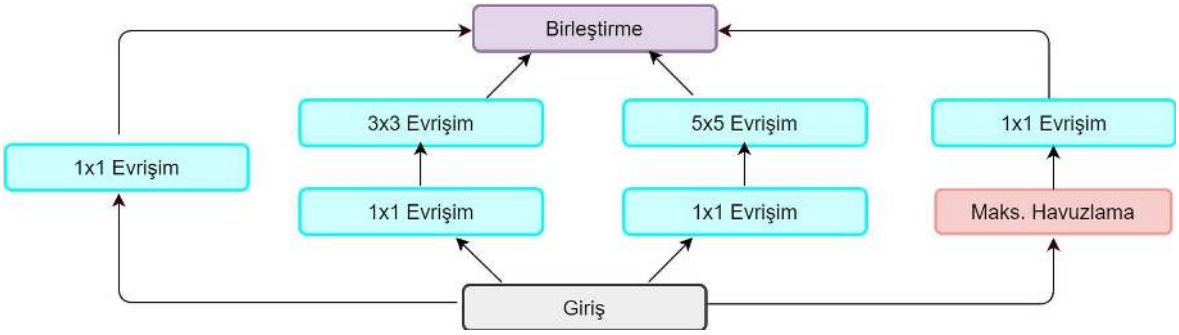


Şekil 2.6: VGG-16 mimarisinin genel yapısı.

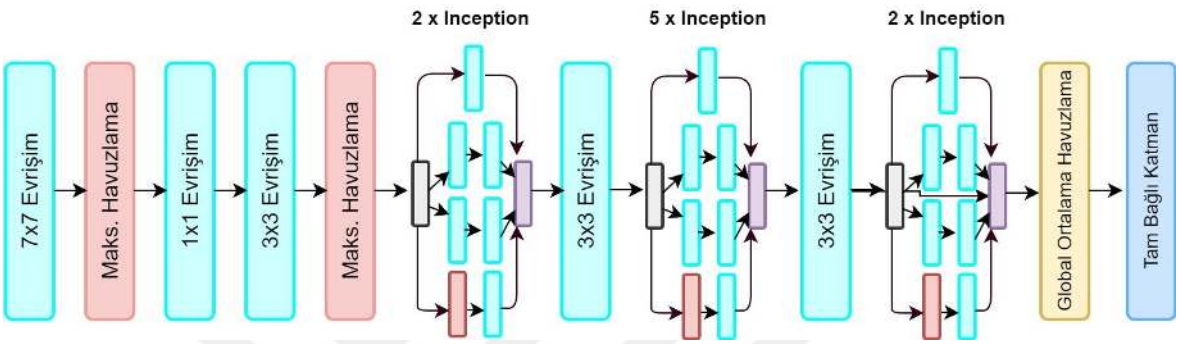
2.2.3 GoogLeNet

GoogLeNet [17] mimarisi, derin evrişimsel sinir ağlarında yüksek temsil gücünü korurken hesaplama maliyetini ve parametre sayısını azaltmayı hedefleyen yenilikçi bir yaklaşım olarak geliştirilmiştir. Bu mimari, ağ performansının yalnızca derinlik ile değil, aynı zamanda ağın genişliği ile de ilişkili olduğunu ortaya koyarak derin öğrenme literatürüne önemli bir katkı sunmuştur. 2014 yılında ImageNet Görsel Tanıma Yarışması'nda (ILSVRC) birincilik elde eden GoogLeNet, VGG gibi derin ancak parametre açısından maliyetli ağlara güçlü bir alternatif oluşturmuştur.

Mimarinin temel taşı olan Inception modülü, geleneksel ardışık katman yapısından farklı olarak, aynı seviyede farklı boyutlardaki filtrelerin (1×1, 3×3 ve 5×5) paralel olarak çalıştırılmasını sağlamaktadır. Bu yaklaşım, ağın görüntü içindeki hem yerel hem de daha geniş ölçekli öznitelikleri eş zamanlı olarak yakalamasına olanak tanımaktadır. Ancak, farklı boyuttaki filtrelerin paralel kullanımı hesaplama maliyetini artırma riski taşıdığından, GoogLeNet darboğaz (bottleneck) katmanları olarak adlandırılan 1×1 evrişim filtrelerini kullanmıştır. Bu 1×1 evrişimler, büyük boyutlu filtreler uygulanmadan önce kanal sayısını azaltarak ağın derinleşmesine rağmen hesaplama yükünün kontrol altında tutulmasını sağlamıştır. Inception modülü ve GoogLeNet mimarisinin genel yapısı sırasıyla Şekil 2.7'de ve Şekil 2.8'de gösterilmiştir.



Şekil 2.7: Inception modülü.

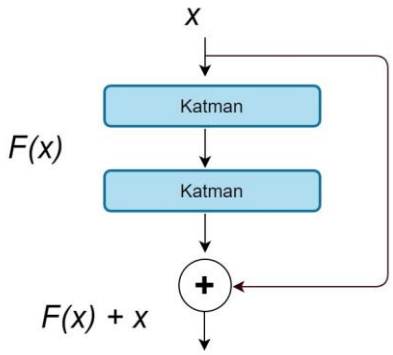


Şekil 2.8: GoogLeNet mimarisinin genel yapısı.

2.2.4 ResNet

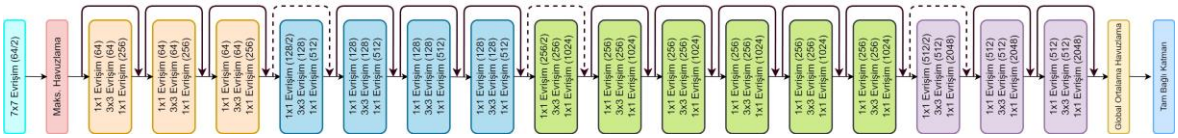
Bilgisayarlı görü çalışmalarında ağ derinliği arttıkça performansın artacağı varsayılsa da, çok derin ağlarda geriye yayılım sırasında gradyanların giderek küçülmesiyle ortaya çıkan gradyan kaybolması problemi, modelin etkin biçimde eğitilmesini zorlaştırmaktadır. Bu sorunu aşmak amacıyla önerilen ResNet(Residual Network) [18] mimarisi, artık (residual) bağlantılar aracılığıyla katmanlar arasında doğrudan bilgi ve gradyan akışı sağlayarak çok daha derin ağların etkin biçimde eğitilebilmesini mümkün kılmıştır. Bu bağlamda ResNet, artık bağlantılar sayesinde derinliğin yol açtığı temel eğitim problemlerini gidererek, modern derin öğrenme mimarilerinde yaygın olarak kullanılan bir yapı haline gelmiştir.

ResNet mimarisinin temel bileşeni olan artık bloklar, katmanların öğrenmesi gereken dönüşümü doğrudan modellemek yerine, giriş ile çıktı arasındaki artık fonksiyonu öğrenmesine dayanmaktadır. Bu yapı, kimlik eşlemesi (identity mapping) üzerinden gerçekleştirilen atlama bağlantıları (skip connections) ile sağlanmaktadır. Matematiksel olarak bu yaklaşım, bir katmanın çıktısını $y = F(x) + x$ şeklinde tanımlamakta olup, gradyanların ağ boyunca daha etkin bir şekilde iletilmesine olanak tanımaktadır. Şekil 2.9'da artık blok gösterimi yapılmıştır.



Şekil 2.9: Artık Blok (Residual Block).

ResNet ailesi, farklı derinlik seviyelerine sahip ResNet-18, ResNet-34, ResNet-50, ResNet-101 ve ResNet-152 gibi varyantlar sunmakta olup, özellikle ResNet-50, temsil gücü ile hesaplama maliyeti arasındaki dengesi nedeniyle literatürde yaygın olarak tercih edilmektedir. ResNet-50 mimarisi, büyük boyutlu filtreli evrişim ve havuzlama katmanları stem bloğu ile başlamakta; bunu izleyen aşamalarda ardışık artık bloklar aracılığıyla hiyerarşik ve giderek daha soyut temsiller öğrenmektedir. Bu aşamalı yapı, derin özelliklerin kademeli olarak inşa edilmesini sağlamaktadır. Şekil 2.10'da Resnet-50 mimarisinin genel yapısı gösterilmiştir.



Şekil 2.10: Resnet-50 mimarisinin genel yapısı.

2.2.5 Transfer Öğrenme

Transfer öğrenme (transfer learning), bir görev üzerinde öğrenilen bilgi ve temsil gücünün, ilişkili başka bir görevin çözümünde yeniden kullanılmasını temel alan bir makine öğrenmesi yaklaşımıdır [19]. Bu yöntemde amaç, daha önce öğrenilmiş bilgi birikiminden faydalanarak yeni bir probleme daha hızlı ve daha etkili şekilde çözüm üretmektir [20]. Geleneksel makine öğrenmesi yöntemlerinde her bir görev için modelin baştan eğitilmesinin gerekmesi, özellikle büyük veri ve yüksek hesaplama maliyetleri açısından önemli bir dezavantaj oluşturmaktadır. Transfer öğrenme ise bu dezavantajları azaltarak eğitim süresini kısaltmakta ve daha az veri ile yüksek performans elde edilmesini mümkün kılmaktadır.

Derin öğrenme alanında transfer öğrenmenin yaygınlaşmasında, büyük ölçekli veri kümeleri üzerinde önceden eğitilmiş ESA modellerinin başarısı önemli bir rol oynamıştır. Özellikle ImageNet gibi geniş ve çeşitli veri setleri kullanılarak eğitilen ESA mimarileri,

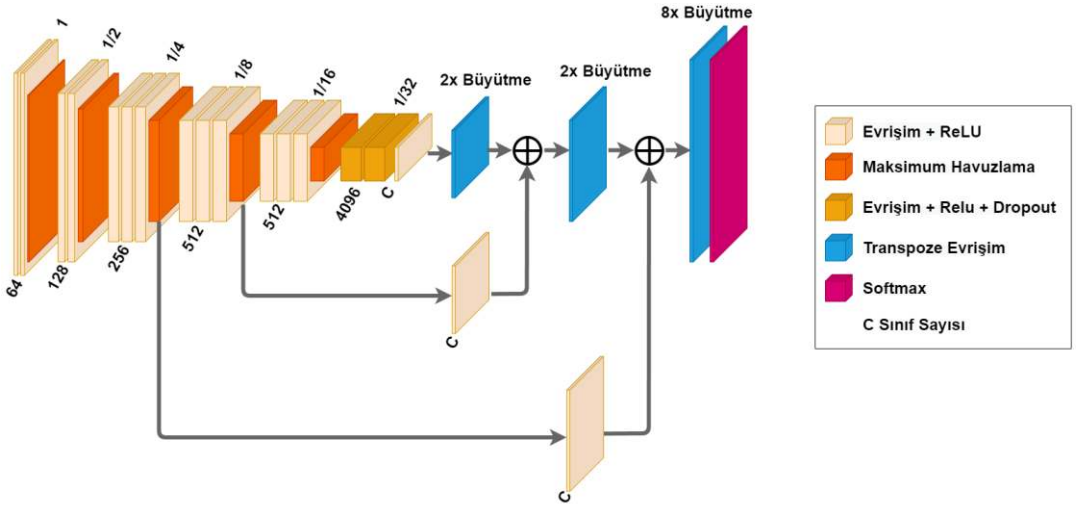
güçlü ve genellenebilir öznitelik temsilleri öğrenebilmekte, bu da söz konusu modellerin farklı problem alanlarında yeniden kullanılmasına olanak sağlamaktadır [21].

Bir probleme yönelik eğitilmiş ESA modellerinin erken katmanlarında öğrenilen temel özniteliklerin, farklı ancak ilişkili görevlerde de geçerliliğini koruması, transfer öğrenmenin temel dayanak noktalarından biridir. Bu bağlamda ön-eğitimli modeller, yeni görevler için güçlü bir başlangıç noktası sunmakta ve derin öğrenme tabanlı çözümlerin daha erişilebilir hale gelmesini sağlamaktadır. Özellikle sınırlı veri ve hesaplama kaynaklarına sahip araştırmacılar için transfer öğrenme, derin öğrenme modellerinin pratik uygulamalarda kullanılabilmesini mümkün kılan etkili bir yöntem olarak öne çıkmaktadır.

2.3 Evrişimsel Sinir Ağları Temelli Görüntü Bölütleme Mimarileri

2.3.1 Tam evrişimli ağlar

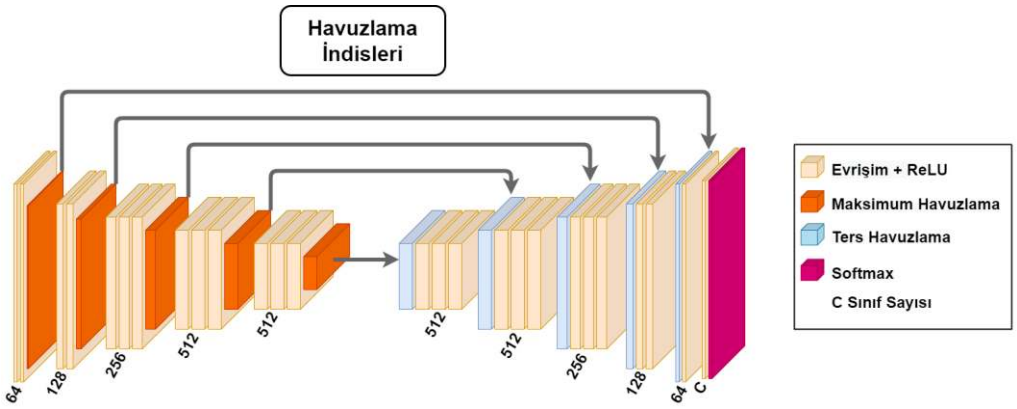
Tam Evrişimli Ağlar (Fully Convolutional Networks - FCN) [22], derin öğrenme tabanlı yöntemlerin semantik bölütleme problemlerine doğrudan uygulanmasını sağlayan ilk yaklaşımlardan biri olarak literatürde önemli bir yere sahiptir. Long ve arkadaşları tarafından önerilen FCN, VGG-16 mimarisindeki tam bağlı katmanların yerine evrişim katmanları kullanarak piksel seviyesinde sınıflandırma gerçekleştirmektedir. Mimarinin tamamen evrişim katmanlarından oluşması, farklı boyutlardaki görüntülerin girdi olarak kullanılabilmesine olanak tanımaktadır. Bu yapıda, art arda uygulanan evrişim ve havuzlama işlemleriyle öznitelik haritalarının uzamsal çözünürlüğü düşürülürken, ağın ilk katmanlarında düşük seviyeli, derin katmanlarında ise yüksek seviyeli öznitelikler öğrenilmektedir. Bölütleme çıktısının girdi görüntüsüyle aynı boyutta olması gerektiğinden, küçülen öznitelik haritaları transpoze evrişim kullanılarak tekrar girdi boyutuna getirilmektedir. FCN mimarisinin ilk örneği olarak önerilen FCN-32s, art arda uygulanan maksimum havuzlama işlemleri sonucunda çözünürlüğü 32 kat azaltmakta ve ardından tek bir transpoze evrişim adımıyla tekrar orijinal boyutuna yükseltmektedir. Ancak bu işlem sırasında uzamsal bilgide kayıplar olmakta ve sonuç olarak kaba bir bölütleme elde edilmektedir. Bu kaybı azaltmak amacıyla, ağın sığ katmanlarında korunan uzamsal bilgi ile derin katmanlarda öğrenilen anlamsal bilgiyi birleştiren atlamalı bağlantılar kullanan FCN-16s ve FCN-8s mimarileri önerilmiştir. Şekil 2.11'de FCN-8s mimarisinin genel yapısı gösterilmiştir.



Şekil 2.11: FCN-8s mimarisinin genel yapısı.

2.3.2 SegNet

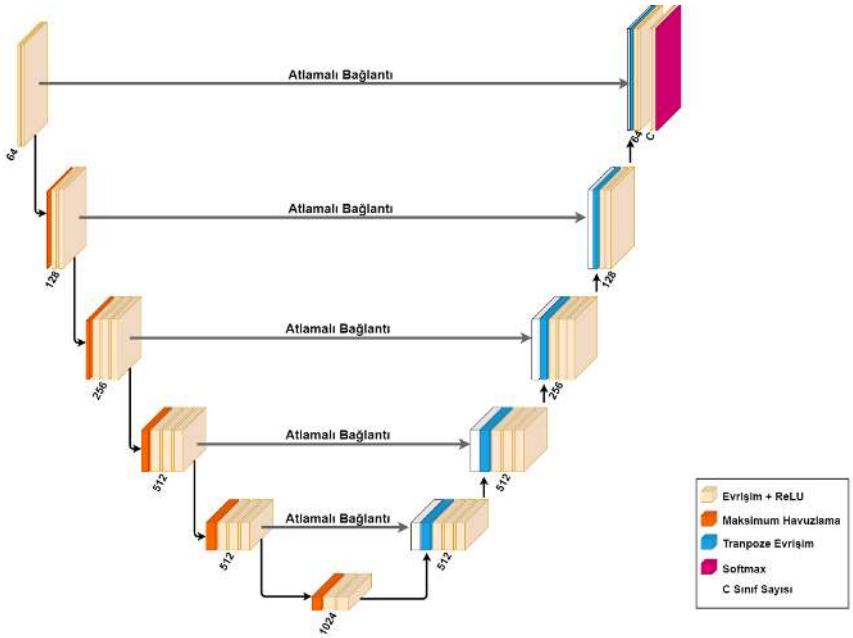
Semantik bölütleme alanında FCN mimarisinin ardından geliştirilen ve özellikle bellek verimliliği ile öne çıkan bir diğer önemli yöntem SegNet [23] mimarisidir. SegNet mimarisi, bir kodlayıcı ağı ve buna simetrik olarak karşılık gelen bir kod çözücü ağından oluşmaktadır. Kodlayıcı ağında tam bağlı katmanları çıkarılmış VGG-16 mimarisi temel alınmıştır. Kodlayıcı bölümde giriş görüntüsüne, sırasıyla evrişim, yığın normalizasyonu, ReLU aktivasyon fonksiyonu ve maksimum havuzlama işlemleri uygulanarak uzamsal çözünürlük kademeli olarak düşürülür; bu sayede görüntüye ait ayırt edici öznitelikler daha düşük çözünürlükte kodlanmaktadır. Kodlanan öznitelik haritaları, kod çözücü ağında evrişim ve ters havuzlama işlemleriyle tekrar giriş görüntüsü boyutuna büyütülmektedir. Son katmanda ise piksel seviyesinde sınıflandırma gerçekleştirilerek nihai bölütleme çıktısı elde edilmektedir. SegNet mimarisinin ayırt edici özelliklerinden biri, kodlayıcı ağında maksimum havuzlama işlemi sırasında elde edilen havuzlama indislerinin saklanması ve bu indislerin kod çözücü ağında gerçekleştirilen ters havuzlama işlemlerinde kullanılmasıdır. Bu yaklaşım sayesinde, çözünürlük artırımı sırasında öznitelik haritalarının uzamsal konum bilgisi korunmakta ve daha tutarlı bölütleme çıktıları elde edilmektedir. Şekil 2.12'de SegNet mimarisinin genel yapısı sunulmuştur.



Şekil 2.12: SegNet mimarisinin genel yapısı.

2.3.3 U-Net

Semantik U-Net, semantik bölütleme alanında özellikle biyomedikal görüntüler için geliştirilmiş ve literatürde yaygın kabul görmüş bir mimaridir [24]. U-Net, SegNet mimarisine benzer şekilde; ayırt edici özniteliklerin çıkarılması için uzamsal çözünürlüğün düşürüldüğü bir kodlayıcı ağ ve öznitelik haritalarının büyütülerek gerçek görüntü boyutuna dönüştürüldüğü bir kod çözücü ağdan oluşmaktadır. Kodlayıcı ağda her maksimum havuzlama işlemi sonrasında öznitelik haritalarının uzamsal çözünürlüğü yarıya düşerken, kanal sayısı iki katına çıkarılmaktadır. Kod çözücü ağda ise transpoze evrişim işlemleri kullanılarak uzamsal çözünürlük iki katına çıkarılırken öznitelik harita sayısı yarıya düşürülmektedir. Kodlayıcı ağ derinleştikçe yüksek seviyeli özniteliklerin öğrenilmesi sağlanırken, maksimum havuzlama işlemi çözünürlüğün azalmasına ve dolayısıyla detaylı uzamsal bilginin kaybolmasına neden olmaktadır. U-Net mimarisinin en ayırt edici özelliği, bu kaybı telafi etmek amacıyla kodlayıcı ağın daha yüksek uzamsal çözünürlüğe sahip öznitelik haritalarını, kod çözücü ağın karşılık gelen seviyeleriyle atlamalı bağlantılar aracılığıyla birleştirmesidir. Şekil 2.13'te U-Net mimarisinin genel yapısı gösterilmiştir.



Şekil 2.13: U-Net mimarisinin genel yapısı.

2.4 Veri Artırma

2.4.1 Veri artırma kavramı ve temel amaçları

Veri artırma (data augmentation), mevcut bir veri kümesinden yeni ve çeşitli örnekler türetmek amacıyla, giriş verisine etiketini koruyacak biçimde çeşitli dönüşümlerin uygulanmasını ifade etmektedir. Bu yaklaşım, özellikle derin öğrenme modellerinin aşırı öğrenme (overfitting) eğilimini azaltmak, genelleme yeteneğini artırmak ve sınırlı veri senaryolarında daha kararlı modeller elde etmek amacıyla yaygın biçimde kullanılmaktadır [25].

Klasik denetimli öğrenme bağlamında veri artırma, çoğunlukla eğitim veri kümesinin istatistiksel çeşitliliğini artıran bir ön-işleme tekniği olarak değerlendirilmiştir. Ancak öz-denetimli öğrenme yaklaşımlarında veri artırma, bu rolün ötesine geçerek öğrenme probleminin ayrılmaz bir parçası haline gelmiştir. Bu bağlamda veri artırma, modelin hangi görsel özelliklere karşı değişmez, hangilerine karşı ayırt edici temsiller öğrenmesi gerektiğini doğrudan belirleyen bir mekanizma olarak işlev görmektedir.

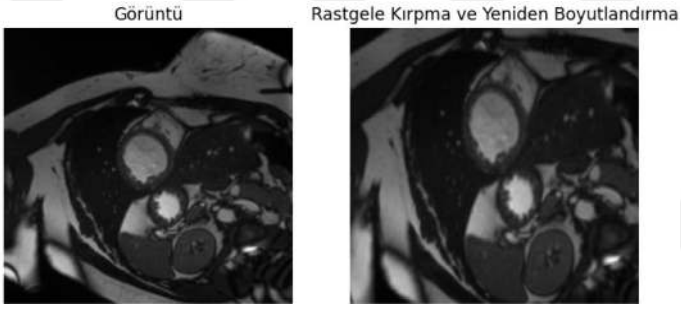
Veri artırma işlemleri sonucunda elde edilen farklı görünümler (views), aynı temel içeriği korumakla birlikte, uzamsal, fotometrik veya yapısal düzeyde belirli farklılıklar içermektedir. Bu farklılıklar sayesinde model, gürültüye ve önemsiz varyasyonlara karşı dayanıklı temsiller öğrenmeye teşvik edilmektedir.

2.4.2 Görüntü tabanlı veri artırma teknikleri

Bu bölümde, literatürde yaygın olarak kullanılan temel görüntü tabanlı veri artırma teknikleri, amaçları ve temsil öğrenme üzerindeki etkileriyle birlikte sunulmaktadır.

2.4.2.1 Rastgele kırpma ve yeniden boyutlandırma

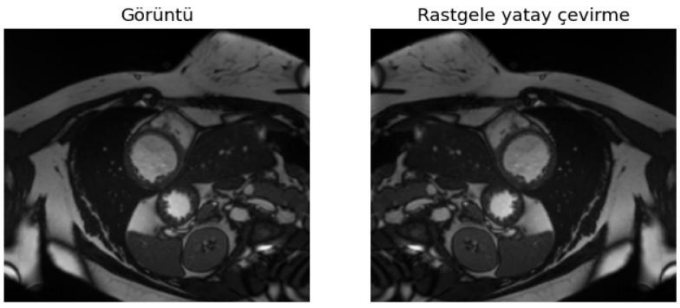
Rastgele kırpma ve yeniden boyutlandırma (Random resized crop), bir görüntünün rastgele seçilen bir alt bölgesinin kırpılarak belirli bir hedef boyuta ölçeklenmesi işlemidir. Bu dönüşüm, modelin nesnelerin görüntü içerisindeki konumuna ve ölçeğine aşırı duyarlı olmasını engelleyerek, uzamsal değişimlere karşı daha dayanıklı temsiller öğrenmesini sağlar. Bu teknik, özellikle nesnenin görüntünün tamamını kaplamadığı durumlarda, modelin yerel ve küresel bağlamı birlikte öğrenmesine katkı sunmaktadır. Şekil 2.14'te rastgele kırpma ve yeniden boyutlandırma uygulanmış görüntü örneği gösterilmiştir.



Şekil 2.14: Rastgele kırpma ve yeniden boyutlandırma uygulanmış örnek görüntü.

2.4.2.2 Rastgele yatay çevirme

Rastgele yatay çevirme (Random horizontal flip), görüntünün yatay eksen boyunca ayna görüntüsünün alınması işlemidir. Görsel simetri içeren veri kümelerinde yaygın olarak kullanılan bu dönüşüm, modelin yön bağımlılığını azaltarak daha genellenebilir temsiller öğrenmesine katkı sağlar. Şekil 2.15'te rastgele yatay çevirme uygulanmış görüntü örneği gösterilmiştir.

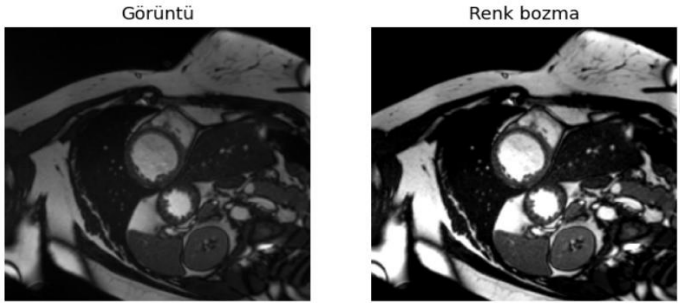


Şekil 2.15: Rastgele yatay çevirme uygulanmış örnek görüntü.

2.4.2.3 Renk bozulmaları

Renk bozulmaları (Color Jitter), görüntünün parlaklık (brightness), kontrast (contrast), doygunluk (saturation) ve renk tonu (hue) gibi fotometrik özelliklerinde rastgele değişiklikler yapılmasını kapsamaktadır. Bu dönüşüm, modelin renk bilgisine aşırı bağımlı olmasını engelleyerek, daha çok şekil, doku ve yapısal özelliklere odaklanmasını teşvik eder. Renk bozulmaları, özellikle aydınlatma koşullarındaki değişkenliğin yüksek olduğu veri kümelerinde temsil öğrenme açısından önemli katkılar sağlamaktadır.

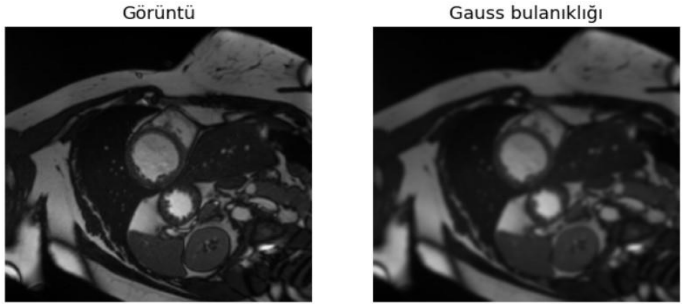
Renk bilgisi içermeyen gri-seviyeli (grayscale) görüntülerde ise doygunluk ve renk tonu bileşenleri anlamsal karşılık taşımadığından, renk bozulmaları yalnızca parlaklık ve kontrast parametreleri ile sınırlandırılmaktadır. Bu bağlamda, gri-seviyeli görüntüler için uygulanan renk bozma işlemleri, yoğunluk tabanlı fotometrik değişimleri temsil etmekte olup, modelin aydınlatma ve kontrast değişimlerine karşı daha dayanıklı temsiller öğrenmesini amaçlamaktadır. Şekil 2.16'da gri-seviyeli bir görüntüye parlaklık ve kontrast temelli renk bozulmaları uygulanmış örnek bir görüntü gösterilmektedir.



Şekil 2.16: Gri-seviyeli bir görüntüye parlaklık ve kontrast temelli renk bozma uygulanmış örnek görüntü.

2.4.2.4 Gauss bulanıklığı

Gauss bulanıklığı (Gaussian blur), görüntüye Gauss kerneli kullanılarak düşük geçiren bir filtre uygulanması işlemidir. Bu dönüşüm, yüksek frekanslı detayları baskılayarak modelin daha küresel ve anlamsal özellikleri öğrenmesine katkı sağlar. Özellikle güçlü veri artırma stratejilerinde, belirli bir olasılıkla uygulanan Gauss bulanıklığı, görünümler arasındaki çeşitliliği artırarak temsil uzayının daha kararlı hale gelmesini desteklemektedir. Şekil 2.17'de Gauss bulanıklığı uygulanmış görüntü örneği gösterilmiştir.



Şekil 2.17: Gauss bulanıklığı uygulanmış örnek görüntü.

2.5 Öz-Denetimli Öğrenme Yöntemleri

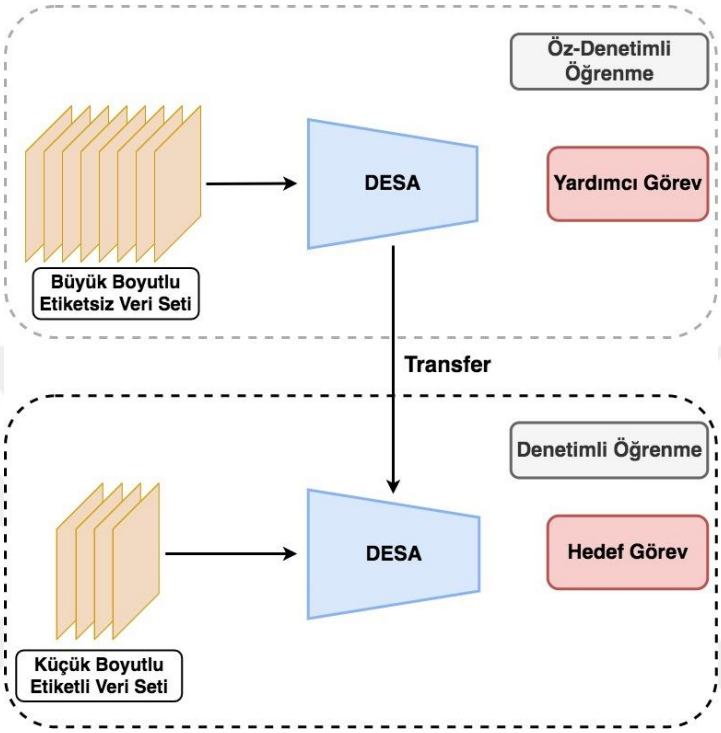
Görsel temsil öğreniminde denetimli öğrenme yöntemleri insan eliyle oluşturulmuş etiketli veriye ihtiyaç duyarken, denetimsiz öğrenme herhangi bir dışsal etiket bilgisi kullanmadan öğrenmeyi gerçekleştirmektedir. Öz-denetimli öğrenme, manuel etiketleme gereksinimini ortadan kaldırması nedeniyle literatürde çoğunlukla denetimsiz öğrenmenin bir alt alanı olarak ele alınmaktadır. [26,27]. Ancak bu iki yaklaşım, öğrenme mekanizması açısından temel bir farklılık gösterir: denetimsiz öğrenme tamamen etiket bağımsız bir süreç izlerken; Öz-denetimli öğrenme, verinin kendi iç yapısından otomatik olarak türetilen sözde etiketler aracılığıyla bir denetim mekanizması kurgulamakta ve tanımlanan yardımcı görevler üzerinden temsil öğrenimini gerçekleştirmektedir [28].

Etiketli veri elde etme sürecinin maliyetli ve yoğun emek gerektiren bir işlem olması, buna karşın internetteki devasa boyutlu etiketsiz verinin kolay erişilebilirliği, Öz-denetimli öğrenme yaklaşımlarına olan ilgiyi artırmıştır [26]. Bu yaklaşım temel olarak yardımcı görev (pretext task) ve hedef görev (downstream task) olmak üzere iki aşamadan oluşur. Görüntü sınıflandırma, nesne tespiti ve bölütleme gibi asıl çözülmek istenen problemler hedef görev olarak tanımlanırken; bu görevlere temel teşkil edecek genelleştirilebilir temsillerin kazanılmasını sağlayan süreçler yardımcı görev olarak adlandırılır [29,30]. Yardımcı görevlerin çözülmesiyle gerçekleştirilen bu temsil öğrenimi süreci, modelin ön-eğitim safhasını oluşturur. Ön-eğitim aşamasında öğrenilen ağ ağırlıkları, hedef görev için modelin başlangıç parametreleri olarak kullanılmakta; böylece model eğitimi rastgele ağırlıklarla başlatılmak yerine ön-eğitimli bir ağ üzerinden gerçekleştirilmektedir [31]. Öz-denetimli öğrenme süreci genel olarak aşağıdaki adımlardan oluşmaktadır:

1. Büyük ölçekli etiketsiz veri kümeleri üzerinde, verinin kendi yapısından yararlanılarak otomatik olarak etiketlerin oluşturulması

2. Oluşturulan sözde etiketler kullanılarak yardımcı görev kapsamında ağın eğitilmesi ve bu süreçte genelleştirilebilir görüntü temsillerinin öğrenilmesi,
3. Öğrenilen temsillerin, sınırlı sayıda etiketli veri içeren hedef göreve aktarılması

Öz-denetimli öğrenme yaklaşımının genel işleyişi Şekil 2.18'de gösterilmektedir.



Şekil 2.18: Öz-denetimli öğrenme yaklaşımının genel işleyişi.

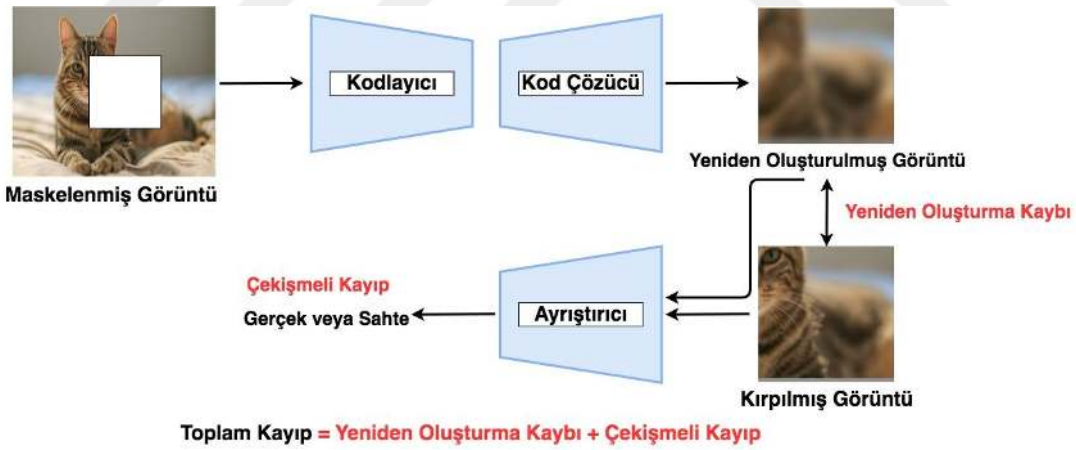
Öz-denetimli öğrenme yaklaşımları, görüntü temsil öğreniminde kullanılan yardımcı görevlerin türüne bağlı olarak beş ana grupta sınıflandırılabilmektedir:

- Yeniden oluşturma tabanlı öz-denetimli öğrenme yöntemleri
- Tahmin tabanlı öz-denetimli öğrenme yöntemleri
- Kümeleme tabanlı öz-denetimli öğrenme yöntemleri
- Karşılaştırmalı öz-denetimli öğrenme yöntemleri
- Karşılaştırmalı olmayan öz-denetimli öğrenme yöntemleri

2.5.1 Yeniden oluşturma tabanlı öz-denetimli öğrenme

Yeniden oluşturma tabanlı öz-denetimli öğrenme yöntemlerinde temel strateji, orijinal görüntü üzerinde belirli deformasyonlar veya değişiklikler yaparak yeni bir girdi oluşturmak ve ardından bu bozulmuş görüntüyü orijinal görüntüye dönüştürebilen bir oto-kodlayıcı mimarisini eğitmektir. Bu yaklaşımın öncü örnekleri olan oto-kodlayıcılar,

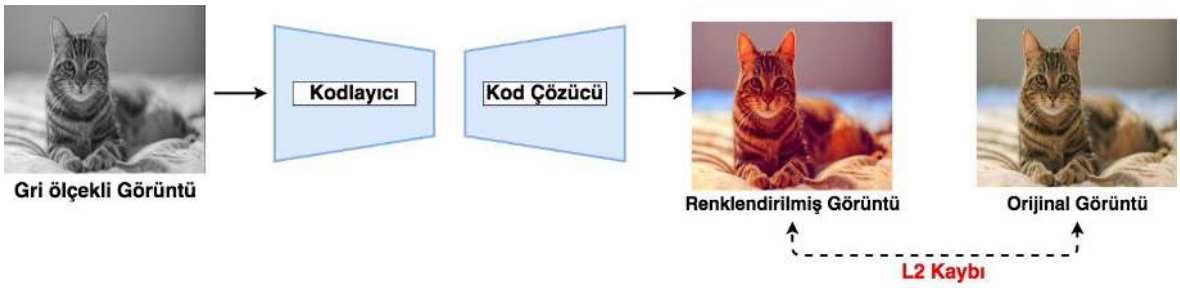
temelde bir kodlayıcı ve bir kod çözücü ağından meydana gelir [31,32]. Bu yapılarda kodlayıcı ağ, girdi görüntüsünü düşük boyutlu bir temsil vektörüne (latent space) indirgerken; kod çözücü ağ ise, bu temsil vektöründen orijinal görüntüyü yeniden oluşturmaya çalışmaktadır. Kodlayıcı ağ maskelenmiş görüntüden görsel temsili öğrenmekte, kod çözücü ağ ise bu temsili kullanarak eksik görüntü bölgesini yeniden oluşturmaktadır. Yeniden oluşturma tabanlı yaklaşımın en yaygın uygulama alanlarından biri, maskelenmiş görüntü bölgelerinin tahmin edilmesi problemidir. Görüntü maskeleme [33] çalışmasında giriş görüntüsünün belirli bölümleri bilinçli olarak gizlenmekte ve bu eksik bölgelerin, görüntünün geri kalan kısımlarından yararlanılarak yeniden üretilmesi amaçlanmaktadır. Bu görev, etiketsiz veriden öğrenme sağlayan kritik bir yardımcı görev işlevi görmektedir. Modelin eğitimi sırasında, eksik bölgelerin genel yapısal özelliklerinin öğrenilmesini teşvik eden yeniden oluşturma kaybı ($L2$) ile üretilen görüntülerin görsel olarak daha gerçekçi olmasını sağlayan çekişmeli kayıp birlikte kullanılmaktadır. Bu yardımcı görevin başarıyla çözülebilmesi için modelin, görüntüde yer alan nesnelerin yapısal özelliklerini, renk dağılımlarını ve anlamsal bileşenlerini öğrenmesi gerekmektedir. Görüntü maskeleme problemine dayalı yeniden oluşturma tabanlı öz-denetimli öğrenme yaklaşımının genel yapısı Şekil 2.19'da gösterilmektedir.



Şekil 2.19: Görüntü maskeleme yönteminin genel yapısı.

Yeniden oluşturma tabanlı öz-denetimli öğrenmenin bir diğer uygulama alanı ise görüntü renklendirme problemidir [34]. Bu yaklaşımda, modele giriş olarak gri ölçekli bir görüntü verilmekte ve ağdan, bu görüntünün renkli karşılığını yeniden üretmesi beklenmektedir. Çalışmada *Lab* renk uzayı kullanılmakta ve modele girdi olarak yalnızca parlaklık bilgisini içeren *L* kanalı verilerek modelden renk bilgisini temsil eden *a* ve *b* kanallarını tahmin etmesi beklenmektedir. Her bir pikselin gerçeğe uygun şekilde renklendirilebilmesi, modelin görüntüdeki nesneleri ayırt edebilmesini ve birbiriyle ilişkili

görsel bölgeleri anlamlı bir şekilde gruplandırabilmesini zorunlu kılmaktadır. Söz konusu çalışmada, öznitelikleri öğrenen kodlayıcı ve renklendirmeyi gerçekleştiren kod çözücü bölümlerinden oluşan tam evrişimsel bir sinir ağı kullanılmaktadır. Modelin eğitimi, tahmin edilen renk değerleri ile orjinal renk değerleri arasındaki farkı minimize eden yeniden oluşturma kaybı ($L2$) kullanılarak gerçekleştirilmektedir. Görüntü renklendirme problemini yardımcı görev olarak kullanan çalışmanın genel yapısı Şekil 2.20'de gösterilmektedir.

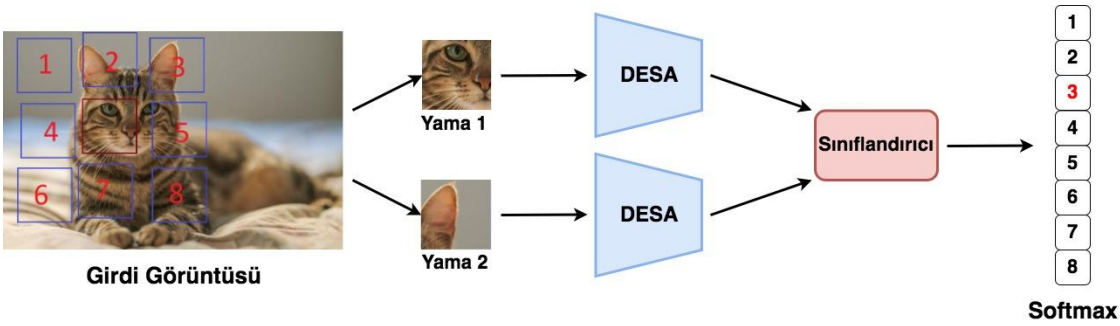


Şekil 2.20: Görüntü renklendirme yönteminin genel yapısı.

2.5.2 Tahmin tabanlı öz-denetimli öğrenme yöntemleri

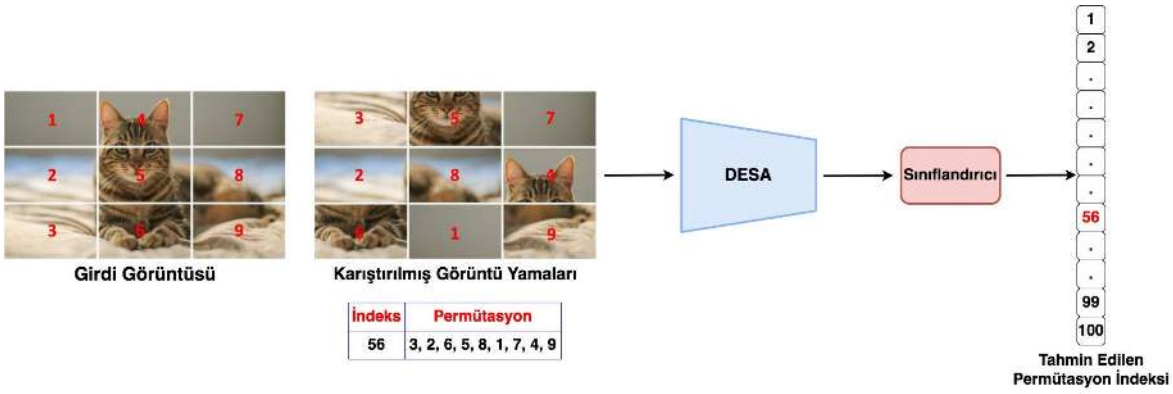
Tahmin temelli öz-denetimli öğrenme yöntemlerinde, verinin sahip olduğu zengin uzamsal bağlamdan faydalanılarak otomatik etiketler oluşturulmaktadır. Türetilen bu etiketleri tahmin etmeye yönelik yardımcı görevler üzerinden eğitilen modeller, manuel etiketlere ihtiyaç duymaksızın genelleştirilebilir görüntü temsillerini öğrenebilmektedir.

Görüntü parçalarının birbirlerine göre konumlarının tahmin edilmesini yardımcı görev olarak ele alan görece pozisyon tahmini yöntemi [35], tahmin tabanlı öz-denetimli öğrenme yaklaşımlarının ilk örnekleri arasında yer almaktadır. Bu görevde amaç, bir görüntüden rastgele seçilen bir yamanın, merkez yama ile olan uzamsal ilişkisini tahmin etmektir. Bu doğrultuda, giriş görüntüleri öncelikle 3×3 boyutunda yamalara bölünmekte ve ardından merkez yama ile ona komşu sekiz yamadan rastgele seçilen bir yama olmak üzere yama çiftleri oluşturulmaktadır. Oluşturulan bu yama çiftleri, AlexNet mimarisinin omurga olarak kullanıldığı ve ağırlıkları paylaşan ikiz Evrişimsel Sinir Ağına (ESA) girdi olarak verilmektedir. Her iki ağdan elde edilen temsilleri birleştirilerek, seçilen yamanın merkez yamaya göre konumu sınıflandırma problemi olarak tahmin edilmektedir. Öğrenme sürecinde modelin düşük seviyeli görsel ipuçlarına odaklanarak basit çözümlere ulaşmasını engellemek amacıyla, yamalar arasında boşluklar bırakılmakta ve yama konumlarında ufak kaydırmalar yapılarak görevin zorluk derecesi artırılmaktadır. Görece pozisyon tahmini yardımcı görevine dayalı öz-denetimli öğrenme yaklaşımının genel yapısı Şekil 2.21'de gösterilmektedir.



Şekil 2.21: Görece pozisyon tahmini yönteminin genel yapısı.

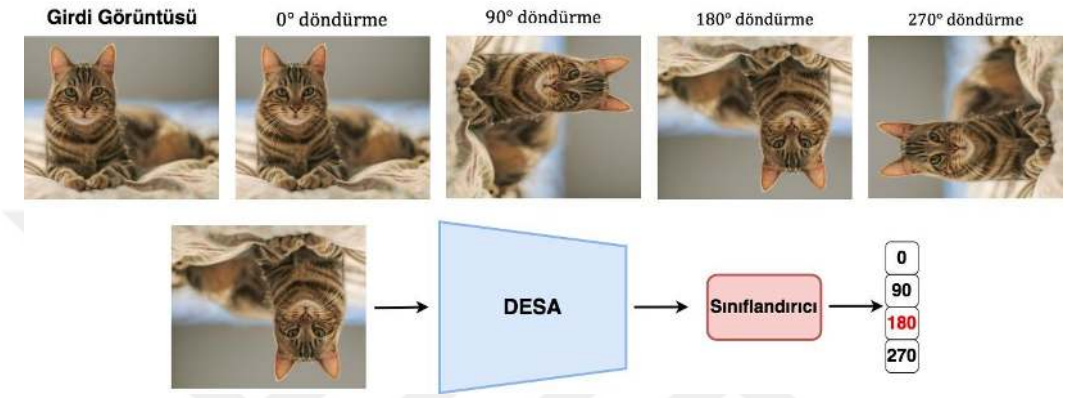
Yapboz çözme [36], tahmin tabanlı öz-denetimli öğrenme kapsamında önerilen bir diğer önemli yardımcı görevdir. Bu yaklaşımda, giriş görüntüsü öncelikle 3×3 boyutunda yamalara bölünmekte ve ardından bu yamalar belirli permütasyonlarla karıştırılarak bir yapboz yapısı oluşturulmaktadır. Model, karıştırılan bu yamaların orijinal konumlarını tahmin edecek şekilde eğitilmekte, böylece modelin görüntüdeki nesneleri tanıması ve nesneleri oluşturan parçalar arasındaki uzamsal ilişkileri öğrenmesi hedeflenmektedir. Görece pozisyon tahmini çalışmasına benzer şekilde, yamalar arasında rastgele boşluklar bırakılarak modelin düşük seviyeli görsel ipuçlarından faydalanıp basit çözümlere ulaşması engellenmektedir. Ayrıca, 9 yama için mevcut olan 362.880 (9!) olası permütasyonun oluşturduğu devasa çözüm uzayını yönetilebilir kılmak adına, eğitim sürecinde bu olasılıkların tamamı yerine, aralarındaki Hamming mesafesi en yüksek olan permütasyonlardan seçilmiş sınırlı bir alt küme kullanılmaktadır. Yardımcı görev olarak yapboz çözme görevinin kullanıldığı tahmin temelli öz-denetimli öğrenme yaklaşımının genel yapısı 2.22 ‘de verilmiştir.



Şekil 2.22: Yapboz çözme yönteminin genel yapısı.

Tahmin tabanlı öz-denetimli öğrenme yöntemlerinden bir diğerinde, giriş görüntüsüne uygulanan dönme açısının tahmin edilmesi yardımcı görev olarak ele alınmaktadır [37]. Bu yöntemde, giriş görüntülerinin 0°, 90°, 180° ve 270° açılarında

döndürülmesiyle oluşturulan görüntüler ve bunlara karşılık gelen döndürme açıları, etiket bilgisi olarak kullanılmaktadır. Oluşturulan veri seti kullanılarak, dört sınıflı bir sınıflandırma problemine benzer biçimde, ESA giriş görüntüsünün hangi açıyla döndürüldüğünü tahmin edecek şekilde eğitilmektedir. Bu yardımcı görevin çözümü, görüntüdeki nesnelerin ve bunlar arasındaki uzamsal ilişkilerin yüksek seviyeli temsillerinin öğrenilmesine dayanmaktadır. Görüntü döndürme açısı tahminine dayalı öz-denetimli öğrenme yönteminin genel yapısı Şekil 2.23'te gösterilmektedir.



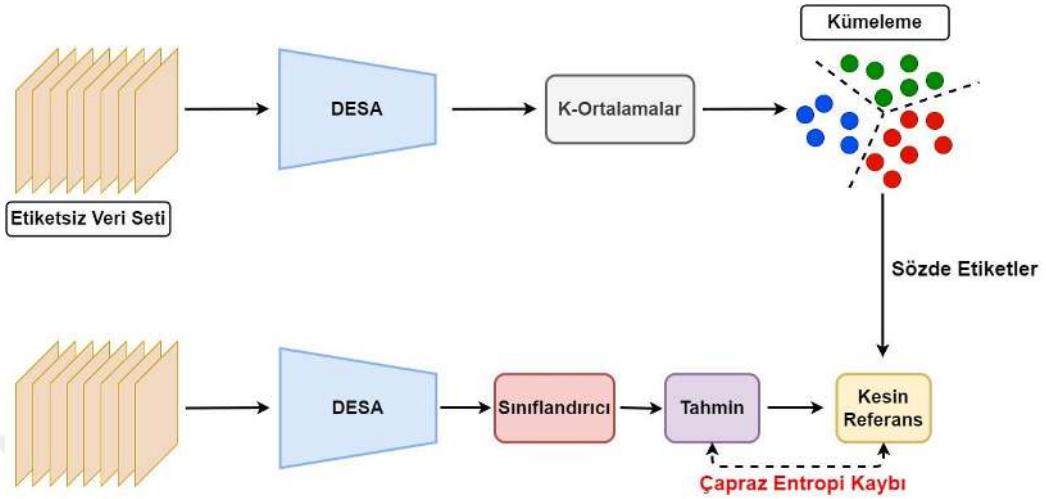
Şekil 2.23: Görüntü döndürme açısı tahmininin genel yapısı.

2.5.3 Kümeleme tabanlı öz-denetimli öğrenme yöntemleri

Kümeleme temelli öz-denetimli öğrenme yöntemleri, etiket bilgisini görüntü temsillerinin kümelenmesi yoluyla otomatik olarak üretilmektedir. Bu yaklaşımlar genel olarak çevrim dışı ve çevrim içi olmak üzere iki temel kategoriye ayrılmaktadır. Çevrim dışı yaklaşımlarda, veri kümesinde yer alan tüm görüntüler ESA'dan en az bir kez geçirilerek elde edilen görüntü temsilleri kümeleme işlemine tabi tutulmakta ve elde edilen bu kümeler eğitim sürecinde sözde etiketler olarak kullanılmaktadır. Buna karşılık, çevrim içi yöntemlerde kümeleme ve temsil öğrenme süreçleri eş zamanlı olarak yürütülmekte, böylece model parametreleri ve küme atamaları birlikte güncellenmektedir.

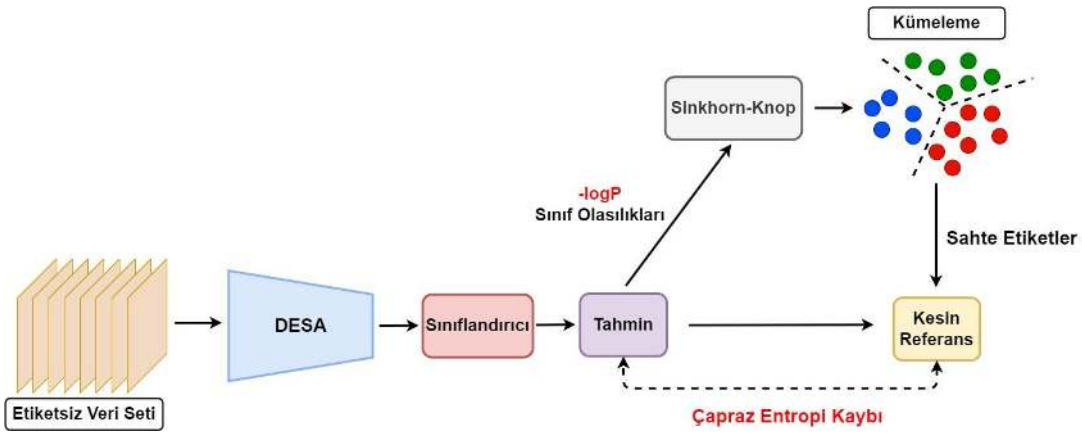
Öz-denetimli öğrenmede kümeleme mekanizmasını kullanan öncü çalışmalardan biri Derin Kümeleme [38] yöntemidir. Bu çalışmada, kümeleme ve öğrenme adımlarının ardışık olarak yürütüldüğü çevrim dışı kümeleme temelli bir öz-denetimli öğrenme yaklaşımı benimsenmektedir. Bu yöntemde, öncelikle girdi görüntüleri bir ESA üzerinden geçirilerek görüntü temsilleri elde edilmektedir. Bu temsillerin ayırt ediciliğini artırmak ve hesaplama maliyetini azaltmak amacıyla sırasıyla temel bileşenler analizi, beyazlatma ve $L2$ normalizasyon işlemleri uygulanmaktadır. Boyutu indirgenmiş ve normalize edilmiş temsiller k-ortalamalar tabanlı kümeleme ile gruplanmakta ve elde edilen sözde etiketler

aracılığıyla ESA eğitimi gerçekleştirilmektedir. Kümeleme ve öğrenme adımları ardışık biçimde yürütülerek daha genellenebilir görüntü temsilleri elde edilmesi amaçlanmaktadır. Derin Kümeleme yaklaşımının genel yapısı Şekil 2.24'te gösterilmektedir.



Şekil 2.24: Derin kümeleme yönteminin genel yapısı.

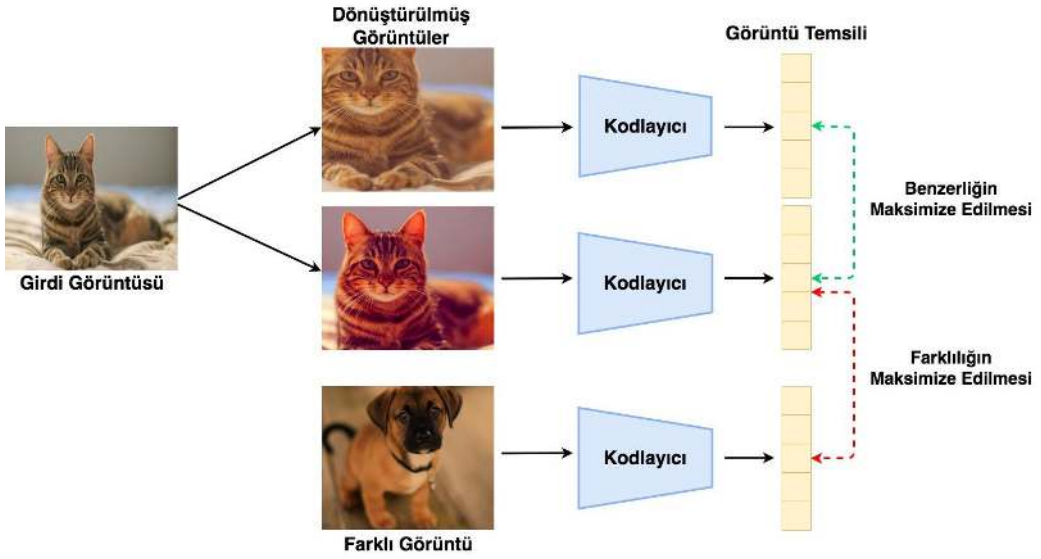
Kümeleme ve öğrenme süreçlerini tek bir optimizasyon çerçevesi altında birleştiren SeLA [39], çevrim içi kümeleme tabanlı öz-denetimli öğrenme yaklaşımlarının önemli bir örneğidir. Bu yaklaşımda, model tarafından üretilen görüntü temsilleri doğrudan öğrenme sürecine dahil edilmekte ve küme atamaları eğitim sırasında dinamik olarak güncellenmektedir. SeLA yöntemi, görüntü temsillerinin kümelere dengeli bir biçimde atanmasını sağlamak amacıyla kümeleme problemini en uygun aktarım problemi olarak ele almaktadır. Sinkhorn–Knopp algoritması [40], ESA çıktılarından türetilen sınıf olasılıklarına dayalı maliyet matrisi ile kümeler arasında dengeli örnek dağılımını zorunlu kılan kısıtı birlikte ele alarak sözde etiketleri üretmektedir. Elde edilen bu sözde etiketler, ağın eğitiminde doğrudan kullanılarak kümeleme ve öğrenme adımlarının eş zamanlı olarak yürütülmesini sağlamaktadır. SeLA yönteminin genel yapısı Şekil 2.25'te gösterilmektedir.



Şekil 2.25: SeLA yönteminin genel yapısı.

2.5.4 Karşılaştırmalı öz-denetimli öğrenme yöntemleri

Karşılaştırmalı öz-denetimli öğrenme, görüntüler arasındaki benzerlik ve farklılıkları karşılaştırarak anlamlı temsiller öğrenmeyi amaçlayan bir yaklaşımdır [41–43]. Bu yaklaşımın temel fikri, benzer görüntülerin temsil uzayında birbirine yaklaştırılması, farklı görüntülerin ise temsil uzayında birbirinden uzaklaştırılmasını sağlayacak bir öğrenme hedefi tanımlamaktır. Görüntüler arası benzerlik ilişkisi, bir görüntüye uygulanan farklı veri artırımlarıyla (data augmentation) türetilen farklı görünümler üzerinden otomatik olarak tanımlanmaktadır. Aynı görüntüye uygulanan farklı veri artırımlarıyla elde edilen görüntü çiftleri pozitif çift olarak tanımlanırken, farklı görüntülere uygulanan veri artırımlarıyla elde edilen görüntü çiftleri negatif çift olarak adlandırılmaktadır. Karşılaştırmalı öğrenme yaklaşımına paralel olarak insan beyni de bir nesneyi tanımlarken onun tüm detaylarını depolamak yerine, yalnızca diğerlerinden ayrışmasını sağlayan ayırt edici özellikleri hafızasında tutmaktadır [44]. Karşılaştırmalı öz-denetimli öğrenmenin genel yapısı Şekil 2.26'da gösterilmektedir.



Şekil 2.26: Karşılaştırmalı öz-denetimli öğrenme yaklaşımının genel yapısı.

Karşılaştırmalı öğrenmenin temel hedefi, x orijinal görüntüyü, bu görüntüye benzeyen x^+ pozitif örneği, x orijinal görüntüsünden farklı x^- negatif örneği ve $sim()$ görüntü benzerlik fonksiyonunu göstermek üzere, $sim\big(f(x), f(x^+)\big) \gg sim\big(f(x), f(x^-)\big)$ koşulunu sağlayan bir f kodlayıcısını öğrenmektir. Bu hedefi gerçekleştirmek için karşılaştırmalı kayıp fonksiyonu kullanılmaktadır. Karşılaştırmalı kayıp fonksiyonu ilk olarak denetimli öğrenme uygulamalarında önerilmiş [45–47], daha sonra öz-denetimli öğrenme bağlamında InstDisc [48] çalışmasında kullanılmıştır. Bu yöntemde, her bir

görüntü örneği ayrı bir sınıf olarak ele alınmakta ve ESA bu sınıfları ayırt edecek şekilde eğitilmektedir. Ancak, her görüntünün ayrı bir sınıf olarak tanımlanması, çıktı uzayını aşırı büyüterek geleneksel softmax fonksiyonunun uygulanmasını hesaplama açısından zorlaştırmaktadır. Bu nedenle InstDisc çalışmasında, bu hesaplama zorluğunu aşmak adına, çok sınıflı sınıflandırma problemini ikili bir sınıflandırma görevine indirgeyen ve Softmax fonksiyonuna etkili bir şekilde yakınsayabilen Noise Contrastive Estimation (NCE) [49] kayıp fonksiyonu kullanılmaktadır.

NCE kayıp fonksiyonunun, çok sayıda negatif örneği dikkate alacak biçimde genişletilmiş bir versiyonu olan InfoNCE [50], karşılaştırmalı öz-denetimli öğrenme yaklaşımlarında yaygın olarak tercih edilen bir kayıp fonksiyonudur. z_i , z^+ aynı görüntünün farklı veri artırımlarından elde edilen görüntülerinin temsilleri, z_i , z^- ise farklı görüntülerin veri artırma dönüşümleri sonrasında karşılık gelen görüntülerinin temsilleri, τ softmax işlemi sonucunda elde edilen olasılık dağılımının keskinliğini kontrol eden sıcaklık katsayısını ve $N-1$ negatif çift sayısını göstermek üzere, InfoNCE kayıp fonksiyonu Denklem 2.6'da verilmiştir.

$$\mathcal{L}_{InfoNCE} = -log \frac{\exp(sim(z_i, z^+)/\tau)}{\exp(sim(z_i, z^+)/\tau) + \sum_{j=1}^{N-1} \exp(sim(z_i, z_j^-)/\tau)} \tag{2.6}$$

Görüntü çiftlerine ait temsillerinin birbirine olan yakınlığını nicel olarak değerlendirmek amacıyla çeşitli benzerlik ölçütleri kullanılmaktadır. Bu ölçütler arasında, iki vektör arasındaki kosinüs açısını temel alan kosinüs benzerliği, görüntü temsillerinin karşılaştırılmasında yaygın olarak tercih edilmektedir. Veri artırımı uygulanan görüntülerin temsilleri z_i ve z_j olmak üzere, görüntü temsilleri arasındaki kosinüs benzerliği $s_{i,j}$, Denklem 2.7'de verilmiştir.

$$s_{i,j} = \frac{z_i^T z_j}{(\| z_i \| \| z_j \|)} \tag{2.7}$$

InfoNCE [52] kayıp fonksiyonu, özünde pozitif çiftler arasındaki karşılıklı bilgiyi (mutual information) maksimize etmeyi amaçlayan bir tahmin mekanizmasıdır. Yüksek boyutlu verilerde karşılıklı bilgiyi doğrudan hesaplamak teknik olarak oldukça zor olduğundan, InfoNCE bu değerin alt sınırını (lower bound) tahmin etmek için kullanılmaktadır. Denklem 2.8'de ifade edildiği üzere, bu alt sınırın yukarı çekilmesi ve dolayısıyla modelin daha ayırt edici temsiller öğrenmesi iki temel faktöre bağlıdır: InfoNCE

kaybının minimize edilmesi ve kullanılan negatif çift sayısının artırılması. Matematiksel bir zorunluluk olarak, negatif örnek sayısı arttıkça karşılıklı bilgi tahmini daha hassas hale gelmekte, bu durum da modelin görüntüdeki gürültüden arınmış ve genelleştirilebilir özellikleri öğrenmesini sağlamaktadır. Bu nedenle, karşılaştırmalı öğrenme yaklaşımlarının başarısı doğrudan kullanılan negatif çiftlerin çokluğuna ve çeşitliliğine dayanmaktadır.

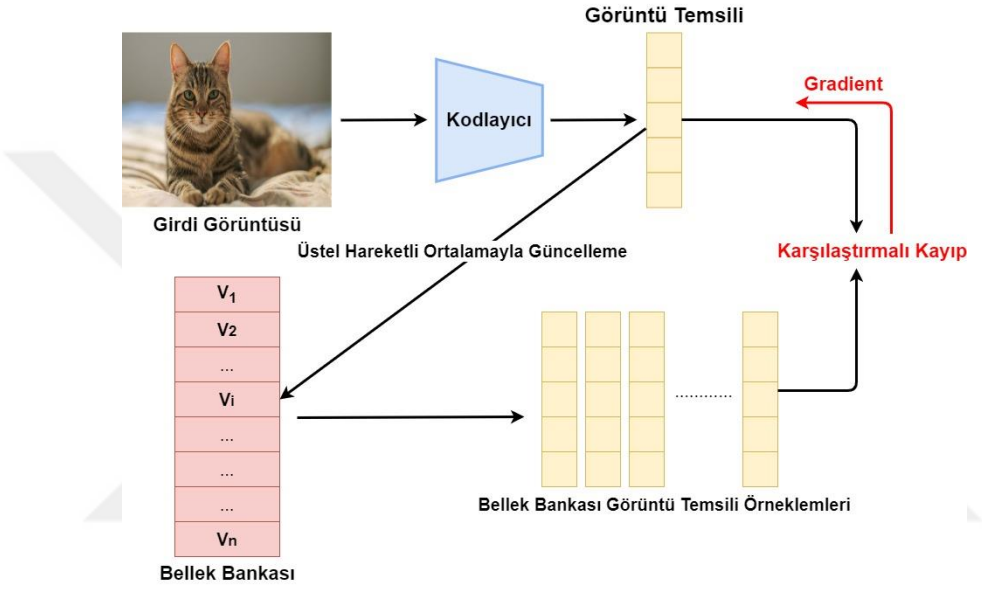
$$I(z_i, z^+) \geq \log(N) - \mathcal{L}_{InfoNCE} \tag{2.8}$$

Exemplar [51], veri kümesindeki her bir görüntünün bağımsız birer sınıf olarak tanımlandığı ve modelin bu sınıfları birbirinden ayırt etmesiyle eğitildiği öncü bir Öz-denetimli öğrenme yaklaşımıdır. Bu yaklaşımda, veri setinden rastgele orijinal bir görüntü seçilmekte ve bu görüntüye çeşitli rastgele görüntü veri artırımları uygulanarak bu görüntünün çok sayıda varyasyonları türetilmektedir. Oluşturulan bu varyasyonların tamamı, orijinal görüntüye ait sözde sınıfa atanmakta ve modelden bu yapay sınıfları birbirinden ayırt edebilecek nitelikte bir sınıflandırma görevi yürütmesi beklenmektedir. Bu sayede model, bir görüntünün farklı görünümlerine karşın temel yapısal özelliklerini ayırt edebilmekte ve veri artırma dönüşümlerine karşı değişmezlik gösteren temsiller oluşturmaktadır. Şekil 2.27'de, seçilen bir görüntü yaması ile bu yamaya uygulanan rastgele veri artırımları sonucunda elde edilen sözde sınıfa ait örnekler gösterilmektedir.

Şekil 2.27: Exemplar yöntemi sözde sınıf örneği.

Exemplar yönteminde, veri kümesinden seçilen her bir görüntü için ayrı bir sözde sınıf tanımlanması, görüntü sayısının artmasıyla birlikte softmax tabanlı sınıflandırmanın hesaplama açısından ölçeklenebilirliğini zorlaştırmaktadır. Bu sınırlılığı gidermek amacıyla InstDisc çalışması, çok sınıflı sınıflandırma problemini ikili sınıflandırma çerçevesine dönüştüren ve softmax fonksiyonuna yakınsayan NCE karşılaştırmalı kayıp fonksiyonunu önermektedir. Bu yaklaşımda her bir görüntü bağımsız bir sınıf olarak ele alınmakta ve ESA,

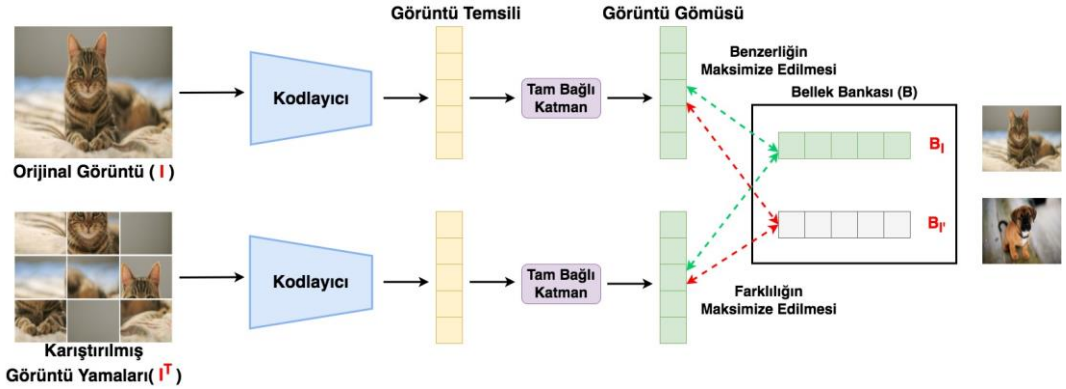
görsel olarak benzer örnekleri temsil uzayında birbirine yakınlaştırırken, farklı örnekleri birbirinden uzaklaştıracak şekilde eğitilmektedir. Çalışmanın getirdiği en önemli yeniliklerden biri, geleneksel parametrik softmax fonksiyonundaki ağırlık vektörleri yerine, görüntü temsillerinin doğrudan birbirleriyle karşılaştırıldığı parametrik olmayan softmax fonksiyonu önermesidir. Ayrıca, hesaplama maliyetini azaltmak amacıyla görüntü temsilleri her iterasyonda yeniden hesaplanmak yerine bir bellek bankasında saklanmakta ve bu temsiller üstel hareketli ortalama yöntemi kullanılarak iteratif olarak güncellenmektedir. InstDisc yönteminin genel yapısı Şekil 2.28'de gösterilmektedir.



Şekil 2.28: InstDisc yönteminin genel yapısı.

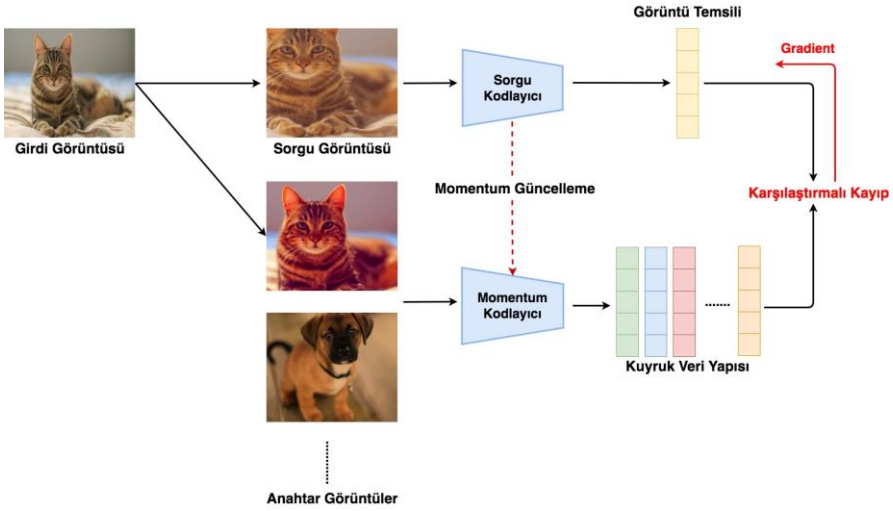
Yardımcı görevler aracılığıyla öznitelik öğrenimi hedefleyen Öz-denetimli öğrenme yöntemleri, yalnızca bu görevlere özgü detaylara odaklandıkları için, öğrenilen temsillerin hedef görevlere genelleştirilmesinde kısıtlılıklar yaşayabilmektedir. Bu sorunu aşmak için önerilen PIRL (Pretext-Invariant Representation Learning) [52], karşılaştırmalı öğrenme prensiplerini geleneksel yardımcı görevlerin sağladığı yapısal bilgilerle birleştirerek, öğrenilen görüntü temsillerinin tanımlanan yardımcı görevlerden bağımsız olmasını hedefleyen hibrit bir Öz-denetimli öğrenme yaklaşımı olarak önerilmiştir. Bu yöntemde, giriş görüntüsü ile aynı görüntünün bir yardımcı görev uygulanmış görünümü, pozitif çifti oluştururken, negatif çift ise giriş görüntüsü ve veri setindeki farklı görüntüler kullanılarak oluşturulmaktadır. Eğitim sürecinde pozitif çiftlerin temsil uzayında birbirine yakınlaşması, negatif çiftlerin ise birbirinden uzaklaşması hedeflenmektedir. Böylece öğrenilen temsillerin, kullanılan yardımcı görevin kendisine değil, görüntünün içeriğine odaklanması amaçlanmaktadır. Karşılaştırmalı öğrenme tabanlı yaklaşım genelleştirilebilir temsil

öğrenimi için çok sayıda negatif örneğe ihtiyaç duyduğundan bu yöntemde InstDisc'e benzer şekilde negatif örnek temsilleri bellek bankasında saklanmaktadır. Belirli bir yardımcı göreve aşırı uyum sağlamayı engellemeyi hedefleyen bu yaklaşım, Öz-Denetimli Öğrenme literatüründe önemli bir yer edinmiş ve daha sonra geliştirilen birçok karşılaştırmalı öğrenme yöntemine kavramsal bir temel oluşturmuştur. Şekil 2.29'da PIRL yönteminin genel yapısı gösterilmiştir.



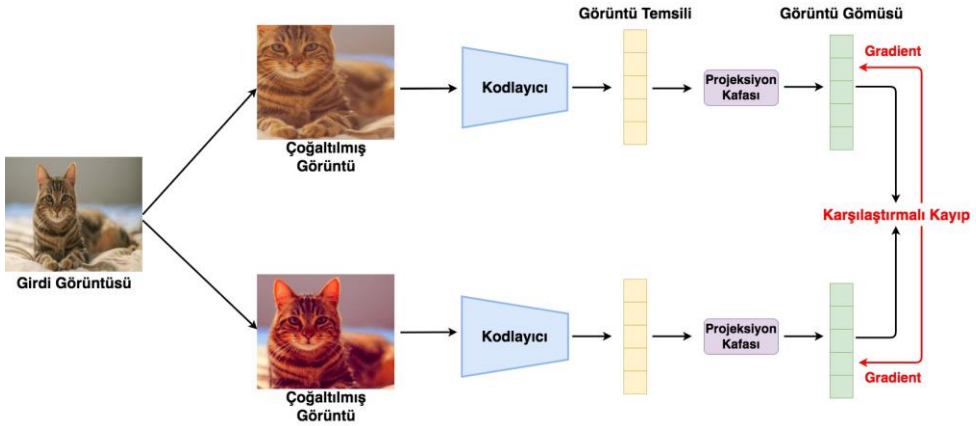
Şekil 2.29: PIRL yönteminin genel yapısı.

Momentum Contrast (MoCo) [42], karşılaştırmalı öz-denetimli öğrenmede genelleştirilebilir temsiller elde edebilmek için gerekli olan çok sayıda negatif örneği, hesaplama maliyeti yüksek büyük mini-yığınlar yerine bir kuyruk veri yapısında tutan bir yaklaşımdır. Bu yaklaşım sorgu kodlayıcı ve momentum kodlayıcı olarak adlandırılan iki kodlayıcıdan oluşmaktadır. Her bir giriş görüntüsüne uygulanan iki farklı veri çoğaltma işlemiyle elde edilen sorgu ve anahtar görünümleri, sırasıyla sorgu kodlayıcı ve momentum kodlayıcıdan geçirilerek görüntü temsilleri üretilmektedir. Kuyruk veri yapısında, anahtar kodlayıcıdan elde edilen güncel görüntü temsilleri her mini-yığın sonrasında kuyruğa eklenirken, güncelliğini yitirmiş temsiller kuyruktan çıkarılmaktadır. Bu süreçte, momentum kodlayıcının ağırlıkları sorgu kodlayıcının ağırlıklarının üstel hareketli ortalaması alınarak yavaş bir biçimde güncellenmekte ve böylece kuyrukta saklanan temsillerin tutarlılığı korunmaktadır. Bu momentum tabanlı güncelleme stratejisi, zaman içerisinde kararlı ve tutarlı bir temsil uzayının oluşmasını sağlamaktadır. Bu sayede model, yüksek donanım gereksinimleri olmaksızın, çok sayıda negatif örnek üzerinden yüksek kaliteli ve genellenebilir görsel temsiller öğrenebilmektedir. Orijinal MoCo mimarisi üzerine inşa edilen MoCo-v2 [53], tasarımına çok katmanlı algılayıcı kafası ve daha güçlü veri çoğaltma yöntemleri eklenerek geliştirilmiştir. Şekil 2.30'da MoCo yönteminin genel yapısı verilmiştir.



Şekil 2.30: MoCo yönteminin genel yapısı.

SimCLR [41], karşılaştırmalı öz-denetimli öğrenmede basit ancak etkili bir çerçeve sunarak, veri çoğaltma stratejilerinin ve büyük mini-yığınların temsil öğrenimi üzerindeki kritik rolünü ortaya koyan bir yöntemdir. Negatif örneklerin saklanmasında, PIRL ve InstDisc bellek bankası ve MoCo kuyruk veri yapısı kullanmakta; SimCLR ise yüksek hesaplama maliyetine sahip büyük boyutlu mini-yığınlardan faydalanmaktadır. SimCLR yaklaşımında, her bir giriş görüntüsüne güçlü ve rastgele veri çoğaltma işlemleri uygulanarak iki farklı görünüm elde edilmekte ve bu görünümler pozitif çifti olarak ele alınmaktadır. SimCLR'ın temel farklılıklarından biri de doğrusal olmayan bir projeksiyon kafasının kullanılmasıdır. Karşılaştırmalı kayıp fonksiyonu, pozitif çiftlere ait gömü vektörlerinin gömü uzayında birbirine yakınlaşmasını teşvik ederken, yığındaki diğer veri artırma yoluyla elde edilmiş görüntülere ait gömü vektörlerinden ayrışmasını sağlamaktadır. SimCLR mimarisi temel alınarak geliştirilen SimCLR-v2 [54] mimarisinde, daha büyük kodlayıcı, daha derin doğrusal olmayan projeksiyon kafası ve daha büyük yığın kullanılmaktadır. Şekil 2.31'de SimCLR yönteminin genel yapısı gösterilmiştir.



Şekil 2.31: SimCLR yönteminin genel yapısı.

SwAV [43], karşılaştırmalı ve çevrim içi kümeleme tabanlı öz-denetimli öğrenme yaklaşımlarının güçlü yönlerini birleştiren ve negatif örnek kullanımına ihtiyaç duymadan temsil öğrenimini gerçekleştiren bir yöntemdir. Çevrim içi kümeleme yaklaşımında, veri setinin tamamının taranmasına gerek duyulmadan, mini-yığınlar bazında üretilen sözde etiketler aracılığıyla ESA eğitilmekte; böylece kümeleme ve öğrenme süreçleri tek bir kayıp fonksiyonu altında yürütülmektedir. SwAV yaklaşımında, veri setini özetleyen ve her biri bir kümeyi temsil eden prototip vektörleri tanımlanmakta, diğer karşılaştırmalı öz-denetimli öğrenme yöntemlerinden farklı olarak görüntü temsilleri doğrudan birbirleriyle karşılaştırılmak yerine, bu temsillerin ortak prototip vektörleri etrafında kümelenmesi sağlanmaktadır. Böylece benzer anlamsal özelliklere sahip görüntülerin aynı kümelerde toplanması hedeflenmektedir. Bu amaç doğrultusunda, her bir giriş görüntüsüne güçlü veri çoğaltma işlemleri uygulanarak birden fazla görünüm elde edilmektedir. SwAV'ın temel yeniliği olan değişmeli atama (swapped assignment) mekanizması ile, bir görüntü görünümüne ait prototip ataması kullanılarak aynı görüntünün diğer görünümüne ait atamanın tahmin edilmesi hedeflenmektedir. Görüntü temsillerinin prototiplere atanmasıyla elde edilen bu küme kodları, ESA'nın eğitimi için gerekli sözde etiketler olarak kullanılmaktadır. Kümeleme süreci, görüntü temsil vektörleri ile prototip matrisinin çarpımından elde edilen maliyet matrisini temel alan optimal taşıma (optimal transport) problemlerinin çözümünde etkin bir yöntem olan Sinkhorn–Knopp algoritması kullanılarak gerçekleştirilmektedir. Algoritma, her bir kümeye dengeli sayıda görüntü temsili atanmasını sağlayan bir kısıt uygulayarak, tüm örneklerin tek bir kümeye çökmesi (collapse) gibi istenmeyen durumların önüne geçmektedir. Bu sayede mini-yığın içerisindeki örnekler prototipler arasında dengeli ve optimize edilmiş bir biçimde dağıtılmaktadır. Bir görüntünün görünümlerinin temsilleri z_s , z_t ve temsillerin sözde etiketleri q_s , q_t olmak üzere, görüntü temsillerinin kayıp fonksiyonu $\ell(z_t, q_s)$, $\ell(z_s, q_t)$ ve toplam kayıp $L(z_t, z_s)$ olarak gösterilmektedir. Denklem (2.9) ve (2.10)'da da görüleceği üzere bir görünümün temsilinin sözde etiketi diğer görünümün temsilleri kullanılarak tahmin edilmeye çalışılmaktadır.

$$I(z_i, z^+) \geq \log(N) - \mathcal{L}_{InfoNCE} \tag{2.9}$$

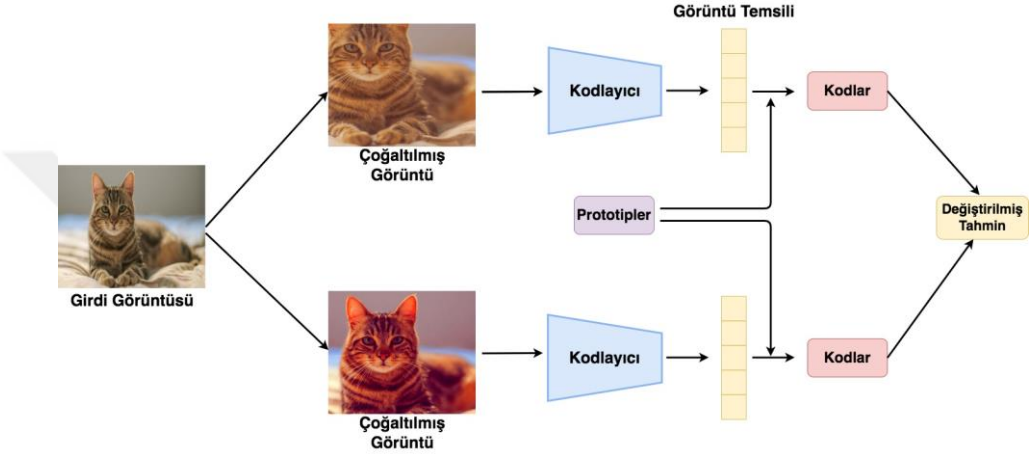
$$\ell(z_t, q_s) = -\sum_k q_s^{(k)} log p_t^{(k)} \tag{2.10}$$

Diğer karşılaştırmalı öğrenme yöntemlerinde doğrudan görüntü temsilleri karşılaştırırken, Denklem (2.11)'de görüleceği üzere bu yöntemde görüntü temsilleri

prototip vektörleri ile karşılaştırılmaktadır. k küme sayısını, τ sıcaklık derecesini ve $p_t^{(k)}$ise görüntü temsilinin k. küme prototipine benzeme olasılığını göstermektedir.

$$p_t^{(k)} = \frac{\exp(\frac{1}{\tau} z_t^T c_k)}{\sum_{k'} \exp(\frac{1}{\tau} z_t^T c_{k'})} \tag{2.11}$$

Şekil 2.32'de SwAV yönteminin değişmeli atama mekanizması ve genel yapısı gösterilmektedir.



Şekil 2.32: SwAV yönteminin genel yapısı.

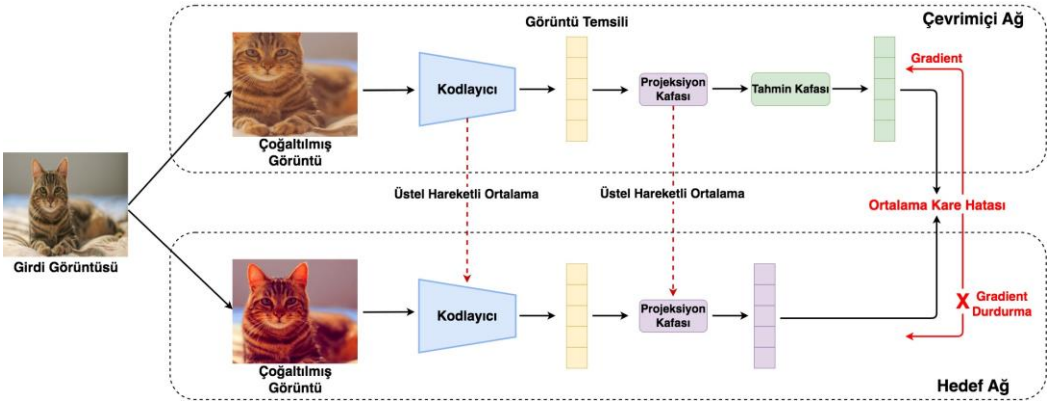
2.5.5 Karşılaştırmalı olmayan öz-denetimli öğrenme yöntemleri

Karşılaştırmalı öğrenme yöntemleri, yüksek kaliteli temsiller elde etmek amacıyla pozitif çiftleri temsil uzayında birbirine yaklaştırırken, negatif çiftleri ise birbirinden uzaklaştırma stratejisine dayanmaktadır. Ancak bu yaklaşım, negatif örneklerin saklanması için bellek bankası veya kuyruk gibi ek veri yapılarına ya da büyük boyutlu mini-yığınlara ihtiyaç duyması nedeniyle yüksek hesaplama maliyetlerini beraberinde getirmektedir. Bu gereksinimler, özellikle sınırlı donanım kaynaklarına sahip ortamlarda ve büyük ölçekli veri kümelerinde eğitim sürecini zorlaştırabilmektedir. Bu sınırlılıkları aşmak amacıyla, son yıllarda negatif örnek kullanımına dayanmayan karşılaştırmalı olmayan (non-contrastive) öz-denetimli öğrenme yaklaşımları geliştirilmiştir. Ancak negatif örneklerin tamamen ortadan kaldırılması, modelin tüm girdiler için benzer veya sabit temsiller üretmesine yol açarak öğrenmenin durduğu model çökmesi (model collapse) problemine neden olabilmektedir. Bu durum, modelin anlamsal ve ayırt edici özellikler öğrenmek yerine, kayıp fonksiyonunu en aza indiren ancak herhangi bir ayırt edici bilgi içermeyen basit çözümlere (trivial solutions) yönelmesiyle sonuçlanmaktadır. Bu nedenle karşılaştırmalı olmayan öz-

denetimli öğrenme yöntemleri, model çöküşünü engellemek amacıyla özel mimari tasarımlar ve öğrenme stratejileri önermektedir.

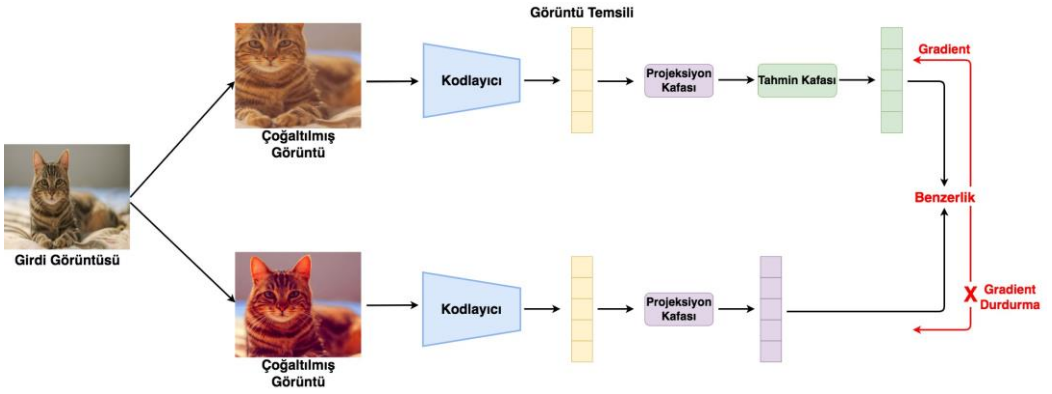
BYOL [55] negatif örnek kullanmadan etkili temsil öğrenimini mümkün kılan ilk karşılaştırmalı olmayan öz-denetimli öğrenme yöntemidir. Bu yaklaşım, model çökmesini engellemek için birbirinden öğrenen çevrim içi ve hedef adlandırılan iki asimetrik ağ yapısı üzerinden temsil öğrenimi gerçekleştiren bir mimariye sahiptir. Çevrim içi ağ bir kodlayıcı, bir projeksiyon kafası ve bu yöntemin en kritik bileşeni olan bir tahmin edici katmanından oluşurken, hedef ağ sadece kodlayıcı ve projeksiyon kafasından oluşmaktadır. Bu yapısal asimetri, modelin tüm girdiler için sabit çıktı üreterek basit çözümlere yönelmesini engelleyen temel mekanizmadır. Eğitim sürecinde, çevrim içi ağın parametreleri geri yayılım yoluyla doğrudan güncellenirken; hedef ağın parametreleri ise çevrim içi ağın ağırlıklarının üstel hareketli ortalamasıyla (EMA) yavaş bir biçimde güncellenmektedir. Bu yaklaşımda, aynı giriş görüntüsüne uygulanan iki farklı veri artırma işlemi sonucunda elde edilen görünümler, çevrim içi ve hedef ağlara girdi olarak verilmektedir. Çevrim içi ağ, hedef ağ tarafından üretilen temsilleri tahmin edecek biçimde optimize edilirken iki ağdan elde edilen temsiller arasındaki farkın azaltılması için karşılaştırmalı kayıp fonksiyonu yerine ortalama kare hatası kullanılmaktadır. Çevrim içi ağdan ve hedef ağdan elde edilen gömü vektörleri sırasıyla $q_\theta(z_\theta)$ ve z'_ξ olmak üzere, öncelikle bu vektörlere $L2$ normalizasyon uygulanmaktadır. Ardından, normalize edilmiş gömü vektörleri arasındaki skaler çarpım hesaplanarak BYOL kayıp fonksiyonu, Denklem (2.12)'de tanımlandığı biçimde elde edilmektedir. Bu yöntemin en dikkat çekici avantajlarından biri ise, geleneksel karşılaştırmalı öğrenme yaklaşımlarının aksine, yığın boyutu ve veri artırma stratejilerindeki değişimlere karşı daha dirençli bir yapı sunmasıdır. Şekil 2.33'te BYOL yönteminin asimetrik ağ yapısı ve momentum tabanlı eğitim süreci gösterilmektedir.

$$\mathcal{L}_\theta^{BYOL} = 2 - 2.\frac{\langle q_\theta(z_\theta), z'_\xi \rangle}{\|q_\theta(z_\theta)\|_2 . \left\|z'_\xi\right\|_2} \tag{2.12}$$



Şekil 2.33: BYOL yönteminin genel yapısı.

Negatif çiftleri kullanmadan temsil öğrenimini gerçekleştiren bir diğer karşılaştırmalı olmayan öz-denetimli öğrenme yaklaşımı SimSiam [56] yöntemidir. BYOL'a benzer şekilde asimetrik bir mimariye sahip olan SimSiam; ağın her iki kolunda bir kodlayıcı ve projeksiyon kafasına sahipken, yapısal asimetriyi sağlamak amacıyla tahmin edici katmanına sadece bir kolda yer vermektedir. SimSiam, model çökmesini önlemek amacıyla mimarisinde asimetri oluşturan ve ağın bir kolunda gradyan akışının durdurulması (stop-gradient) mekanizmasını kullanan, momentum kodlayıcı içermeyen bir BYOL varyantı olarak değerlendirilebilir. Bu yaklaşımda, tek bir görüntüden veri artırımı yoluyla türetilen iki farklı görünüm girdi olarak kullanılmakta ve bu görünümlere karşılık gelen temsiller arasındaki benzerliğin maksimize edilmesi hedeflenmektedir. Şekil 2.34'te SimSiam yönteminin gradyan durdurma prensibine dayalı genel yapısı gösterilmektedir.



Şekil 2.34: SimSiam yönteminin genel yapısı.

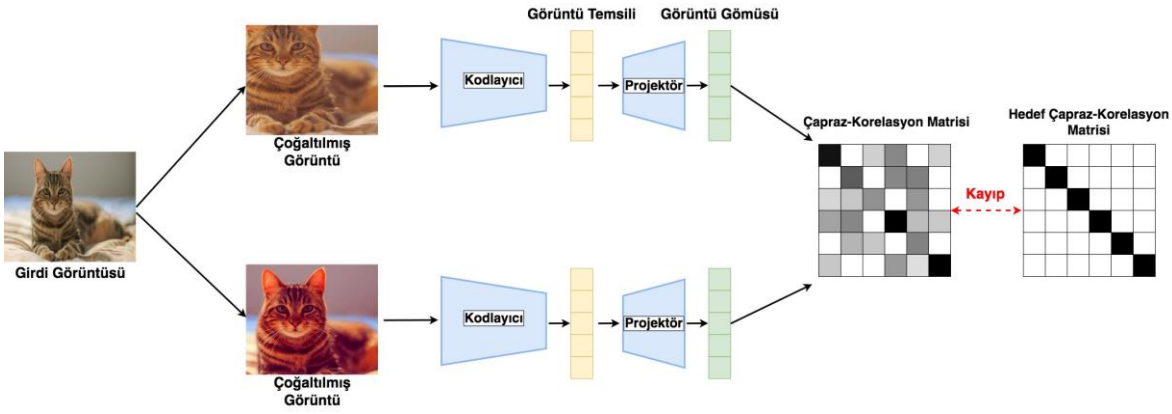
Barlow Twins [57], negatif örnek kullanımına dayanmayan karşılaştırmalı olmayan öz-denetimli öğrenme yaklaşımlarından biri olup, temsil öğrenimini istatistiksel bağımlılıkların azaltılması prensibi üzerine inşa etmektedir. Bu yöntem, BYOL ve SimSiam gibi yaklaşımlardan farklı olarak mimari asimetri, momentum kodlayıcı veya gradient durdurma mekanizmalarına ihtiyaç duymadan model çökmesi sorununu önleyebilmektedir.

Barlow Twins'in temel amacı, aynı görüntünün farklı veri artırmalarıyla elde edilen görünümlerinden çıkarılan temsillerin benzerliğini artırırken, temsil boyutları arasındaki gereksiz bilgi tekrarını (redundancy) en aza indirmektir. Bu yaklaşımda, aynı giriş görüntüsüne uygulanan iki farklı veri artırımı sonucunda elde edilen görünümler, ağırlıkları paylaşılan iki özdeş ağ üzerinden geçirilerek z^A ve z^B görüntü temsil matrisleri elde edilmektedir. Elde edilen bu temsillerin istatistiksel ilişkisini modellemek amacıyla, iki görünüm arasındaki çapraz-korelasyon matrisi hesaplanmaktadır. C çapraz-korelasyon matrisi, b yığını, i ve j görüntü temsillerinin boyut indislerini temsil etmek üzere Denklem 2.13'te verilmiştir.

$$C_{ij} = \frac{\sum_b z_{b,i}^A z_{b,j}^B}{\sqrt{\sum_b \left(z_{b,i}^A\right)^2} \sqrt{\sum_b \left(z_{b,j}^B\right)^2}} \tag{2.13}$$

Yöntemde, elde edilen çapraz korelasyon matrisini birim matrise yaklaştırmayı hedefleyen bir kayıp fonksiyonu önerilmektedir. Böylece köşegen elemanların bire yaklaştırılmasıyla iki görünümün aynı temsil boyutlarında benzerlik göstermesi sağlanırken, köşegen dışı elemanların sıfıra yaklaştırılmasıyla farklı temsil boyutları arasındaki bağımlılıklar ve fazlalık bilgi azaltılmaktadır. Bu kayıp fonksiyonu, bir yandan temsiller arasında değişmezliği teşvik ederken, diğer yandan temsil boyutları arasındaki fazlalığı azaltarak ayırt edici ve bilgi açısından zengin temsillerin öğrenilmesini sağlamaktadır. Denklem 2.14'te, Barlow Twins yönteminde kullanılan kayıp fonksiyonu verilmiştir ve Şekil 2.35'te bu yöntemin genel yapısı gösterilmiştir.

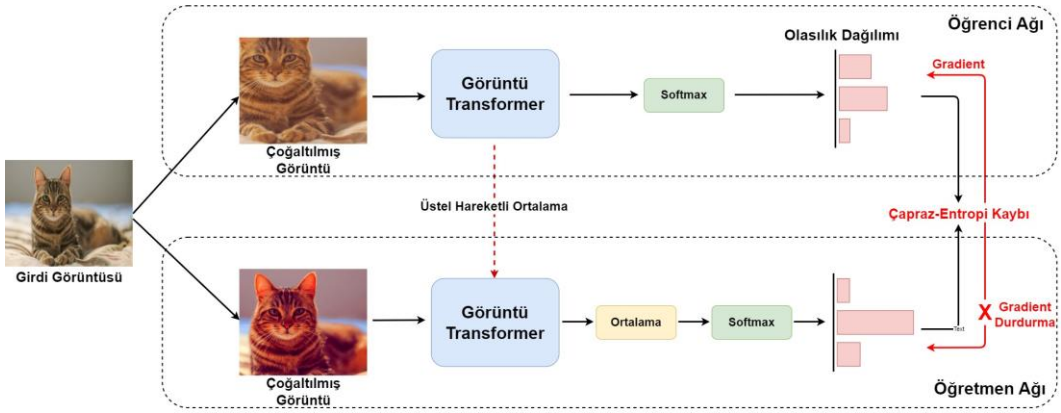
$$\mathcal{L}_{BT} = \sum_i (1 - C_{ii})^2 + \lambda \sum_i \sum_{j \neq i} {C_{ij}}^2 \tag{2.14}$$



Şekil 2.35: Barlow Twins yönteminin genel yapısı.

DINO [58], negatif örnek kullanımına dayanmayan ve öz-denetimli öğrenmeyi kendi kendine bilgi damıtımı (self-distillation) prensibi üzerinden gerçekleştiren karşılaştırmalı olmayan bir temsil öğrenme yaklaşımıdır. Görsel Transformer (Vision Transformer -ViT) mimarilerinin etiketlenmemiş büyük veri kümeleri üzerinde verimli bir şekilde eğitilmesine olanak tanıyan bu yöntem; parametreleri geri yayılım ile güncellenen bir öğrenci ağı ve ağırlıkları öğrenci ağın üstel hareketli ortalaması yoluyla güncellenen bir öğretmen ağından oluşmaktadır. Yöntemin temel başarısı, çoklu ölçekli veri artırımı stratejisi kullanılarak görüntünün yerel parçalarından global bağlamın tahmin edilmesine dayanmaktadır. Bu kapsamda, öğrenci ağı girdi olarak hem görüntünün tamamını kapsayan global görünümleri hem de görüntünün küçük bir bölümünü içeren yerel görünümleri girdi olarak alırken, öğretmen ağı yalnızca görüntünün büyük bir kısmını kapsayan global görünümleri girdi olarak almaktadır. Modelin tüm girdiler için aynı çıktıyı üreterek çökmesini engellemek amacıyla, DINO'da öğretmen ağın çıktılarına softmax fonksiyonu öncesinde merkezleme ve keskinleştirme işlemleri uygulanmaktadır. Merkezleme işlemi, çıktıların belirli sınıflara yığılmasını önleyerek temsil çeşitliliğini artırırken; keskinleştirme işlemi, olasılık dağılımlarının daha belirgin hale gelmesini sağlayarak belirsizliği azaltmaktadır. Öğrenci ağın çıktısına ise doğrudan softmax fonksiyonu uygulanmakta ve öğrenci ağ, öğretmen ağın çıktılarıyla tutarlı tahminler üretmeye zorlayan bir çapraz entropi kayıp fonksiyonu ile eğitilmektedir. DINO sayesinde Vision Transformer mimarileri, herhangi bir nesne etiketi kullanılmaksızın nesne bütünlüğünü yakalayabilen ve sahne içerisindeki nesneleri anlamsal olarak ayırt edebilen güçlü temsil yapıları öğrenebilmektedir. DINO mimarisini temel alan ve ViT [59] yapısını kullanan DINOv2 [60], herhangi bir ince ayar (fine-tuning) gerektirmeksizin; sınıflandırma, bölütleme, nesne tespiti ve derinlik tahmini gibi çeşitli bilgisayarlı görü görevlerinde üstün performans sergilemiştir. DINO yönteminin ön-eğitim süreci yalnızca ImageNet [61] veri seti ile gerçekleştirilirken, DINOv2 için farklı veri setleri ve web kaynaklarından derlenen, toplamda 142 milyon görüntü içeren çok daha büyük ölçekli bir ön-eğitim veri seti (LVD-142M) kullanılmıştır. Öğrenci ve öğretmen ağlarının olasılık dağılımları sırasıyla P_s ve P_t , öğrenci ağının parametreleri θ_s, ve $H(a,b) = -a.logb$ olmak üzere Denklem 2.15'te DINO kayıp fonksiyonu verilmiştir ve Şekil 2.36'da DINO yönteminin genel işleyişi gösterilmiştir.

$$min_{\theta_s} H\big(P_t(x), P_s(x)\big) \tag{2.15}$$



Şekil 2.36: DINO yönteminin genel yapısı.

2.6 Dalgacık Saçılım Ağları

2.6.1 Teorik temeller ve öznitelik çıkarım mekanizması

Dalgacık Saçılım Ağları (Wavelet Scattering Networks) [62,63], ilk olarak Mallat tarafından önerilen, matematiksel temellere dayalı öznitelik çıkarıcılardır. Veriden filtreleri öğrenen geleneksel ESA'ların aksine saçılım dönüşümünde, önceden tanımlanmış sabit bir dalgacık filtre bankası kullanılmaktadır. Bu yapı, dalgacık evrişimleri, doğrusal olmayan modül işlemleri ve düşük geçiren filtreleme operasyonlarının ardışık kullanımıyla, yerel ötelemelere karşı değişmez ve küçük deformasyonlara karşı kararlı hiyerarşik öznitelikler üretmektedir. Dalgacık Saçılım Ağları (DSA), temel yapılarının ötesinde, Sifre ve Mallat [64] tarafından önerilen döndürme-öteleme (roto-translation) saçılım dönüşümü gibi genişletmeler sayesinde, rotasyonlara karşı değişmezlik özelliğini de temsil yeteneklerine dahil edebilmektedir. Şekil 2.37'de DSA'da saçılım katsayılarının elde edilmesinde kullanılan operasyonlar gösterilmiştir.

Şekil 2.37: Dalgacık Saçılım Ağlarında saçılım katsayılarının elde edilmesinde kullanılan temel işlemler.

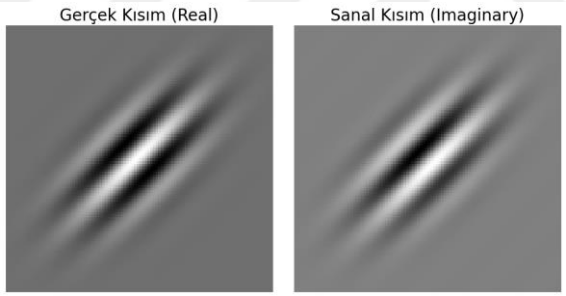
2.6.2 Morlet dalgacığı ve filtre bankası

DSA'da ana dalgacık olarak genellikle Morlet dalgacıkları kullanılmaktadır. Morlet dalgacığı, hem uzamsal hem de frekans alanında iyi yerelleşme özellikleri sunması nedeniyle, görüntülerdeki kenar, doku ve yönelimli yapıları etkin bir şekilde temsil edebilen bir ana dalgacıktır. Ayrıca, frekans alanında dar bantlı bir yapı sunması, farklı ölçek ve

yönelimlerdeki yapısal bileşenlerin ayrıştırılmasını mümkün kılmaktadır. İki boyutlu Morlet dalgacığı, Gauss penceresi ile modüle edilmiş karmaşık bir sinüs dalgası olarak tanımlanmaktadır ve ξ merkez frekansını, σ Gauss penceresinin genişliğini, C_1 ve C_2 normalizasyon katsayılarını ve dalgacığın ortalamasının sıfır olmasını sağlayan düzeltme terimlerini göstermek üzere bir $\psi(u)$ ana dalgacığı şu şekilde ifade edilir:

$$\psi(u) = C_1\left(e^{iu\cdot\xi} - C_2\right)e^{-\frac{|u|^2}{2\sigma^2}} \tag{2.16}$$

Morlet dalgacığı, gerçek ve sanal bileşenlerden oluşmakta olup, bu bileşenler uzamsal alanda farklı faz bilgilerini temsil etmektedir. Gerçek ve sanal bileşenlerin birlikte kullanılması, genlik bilgisinin fazdan bağımsız olarak yakalanmasına imkan tanıyarak, yönelim-duyarlı ve kararlı özniteliklerin elde edilmesini mümkün kılmaktadır. Bu bağlamda, Morlet dalgacığı özellikle yönelimli kenar ve doku yapılarının temsilinde etkin bir rol oynamaktadır. İki boyutlu Morlet dalgacığının gerçek ve sanal bileşenlerine ilişkin örnek görsel Şekil 2.38'de sunulmaktadır.



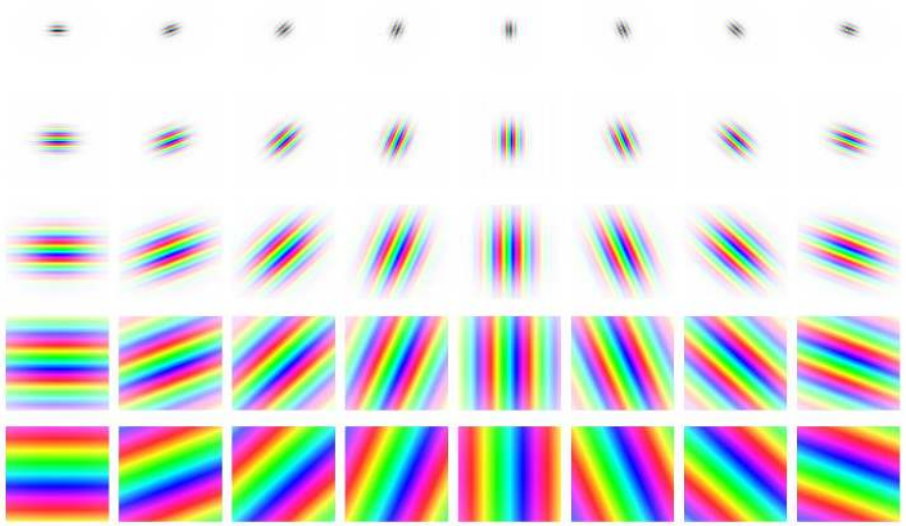
Şekil 2.38: Morlet dalgacığının gerçek ve sanal bileşeni.

Morlet ana dalgacığı, ölçekleme ve döndürme işlemleriyle genişletilerek çok-ölçekli ve çok-yönelimli bir wavelet filtre bankası oluşturulmaktadır. Bu filtre bankası, giriş görüntüsündeki yapısal bileşenlerin farklı çözünürlük seviyelerinde ve farklı açılarda ayrıştırılmasına olanak tanımaktadır. Böylece hem düşük frekanslı global yapılar hem de yüksek frekanslı yerel ayrıntılar birlikte temsil edilebilmektedir. Bu kapsamda, Morlet dalgacığı aşağıdaki şekilde parametrelendirilmiş bir filtre ailesine dönüştürülmektedir:

$$\psi_{j,\ell}(u) = 2^{-2j}\psi\left(2^{-j}r_\ell^{-1}u\right) \tag{2.17}$$

Burada $j \in \{0, \dots, J-1\}$ ölçek indeksini, $\ell \in \{0, \dots, L-1\}$ yönelim indeksini ve R_ℓ dalgacığın ℓ'inci yönelime karşılık gelen döndürme operatörünü ifade etmektedir. Ölçek parametresi J, dalgacıkların uzamsal destek boyutlarını belirleyerek görüntünün farklı

çözünürlük seviyelerinde analiz edilmesini sağlarken; yönelim sayısı L, kenar ve doku gibi yönelim-duyarlı yapıların etkin biçimde yakalanmasına olanak tanımaktadır. Bu şekilde oluşturulan wavelet filtre bankası, frekans alanında logaritmik olarak örneklenmiş bantlara karşılık gelmekte olup, düşük frekanslardan yüksek frekanslara kadar geniş bir spektral kapsama alanı sunmaktadır. Her bir ölçek–yönelim çifti, görüntünün belirli bir frekans bandındaki ve belirli bir yöndeki yapısal bilgisini temsil etmektedir. Bu özellik, saçılım dönüşümünün çok-ölçekli ve yönelim-duyarlı öznitelikler üretmesinin temelini oluşturmaktadır. Örneğin, $J = 5$ ölçek ve $L = 8$ yönelim için oluşturulan Morlet tabanlı dalgacık filtre bankası, toplamda $J \times L$ adet bant-geçiren filtreden oluşmakta olup, bu filtrelerin uzamsal ve frekans alanındaki dağılımı Şekil 2.39'da gösterilmektedir. Şekilde, dalgacıkların genlik bilgisi parlaklık/kontrast ile, faz bilgisi ise renk (hue) ile kodlanmıştır. İlgili filtre bankası, frekans düzleminde farklı yönelimlerde ve farklı bant genişliklerinde yerleşerek, görüntüdeki çok-ölçekli yapısal örüntülerin etkin biçimde ayrıştırılmasını sağlamaktadır.



Şekil 2.39: 5 ölçek ve 8 yönelimli 2 Boyutlu Morlet dalgacık filtre bankası.

2.6.3 Saçılım katsayılarının hesaplanması

Dalgacık, hem uzamsal hem de frekans alanında yerelleştirilmiş ve ortalaması sıfır olan bir fonksiyon olarak tanımlanmaktadır. Saçılım dönüşümünde ölçek sayısı (J) ve yönelim sayısı (L) ile parametrelendirilmiş bir dalgacık filtre bankası uygulanmaktadır. Her bir dalgacık $\psi_{j,\theta}$ ana dalgacığın ψ ölçek 2^j faktörüyle genişletilmesi θ açısıyla döndürülmesiyle oluşturulur; bu da sinyaldeki değişimlerin farklı çözünürlük ve yönlerde yakalanmasını sağlamaktadır. Uzamsal konumu u ile ifade edilen bir $x(u)$ giriş sinyali ele alındığında; bu dönüşüm, sırasıyla sıfırıncı, birinci ve ikinci derece özniteliklere karşılık

gelen S_0x, S_1x ve S_2x saçılım katsayılarını üretmektedir. Sıfırıncı derece saçılım katsayılarının hesaplanması için, giriş gri-seviye görüntüsü x, uzamsal pencere boyutu 2^j ölçeğine karşılık gelen ve genellikle Gauss kerneli olarak seçilen bir düşük geçiren ortalama filtresi ile evrişime tabi tutulmaktadır. Bu işlem aşağıdaki şekilde tanımlanmaktadır:

$$S_0x = x * \phi_J \tag{2.18}$$

Düşük geçiren filtreleme adımı, görüntünün genel yapısına ilişkin kaba bir temsil elde edilmesini sağlamakta ve uzamsal çözünürlüğün 2^j faktörü oranında düşürülmesiyle sonuçlanmaktadır. Sonuç olarak, $N \times N$ boyutundaki bir giriş gri-seviye görüntüye karşılık olarak elde edilen sıfırıncı derece saçılım çıktısı S_0x, $N/2^J \times N/2^J$ uzamsal boyutlarına sahip, çözünürlüğü azaltılmış bir öznitelik haritası üretmektedir. Her ne kadar bu süreç, 2^j ölçeğinden daha küçük yer değiştirmelere karşı değişmezlik sağlasa da, ince ayrıntıların ayırt edilmesi açısından önemli olan yüksek frekans bileşenlerinin kaybolmasına neden olmaktadır. Saçılım dönüşümünün sonraki aşamaları, ardışık dalgacık filtreleme ve doğrusal olmayan işlemler aracılığıyla kaybolan yüksek frekanslı içeriği geri kazanarak bu kısıtlamayı gidermeyi amaçlamaktadır. Birinci derece katsayılar; x sinyalinin karmaşık dalgacıklar ψ_{λ_1} ile evriştirilmesi, ardından bir doğrusal olmayan modül işleminin uygulanması ve son olarak düşük geçiren filtreden geçirilmesiyle hesaplanmaktadır.

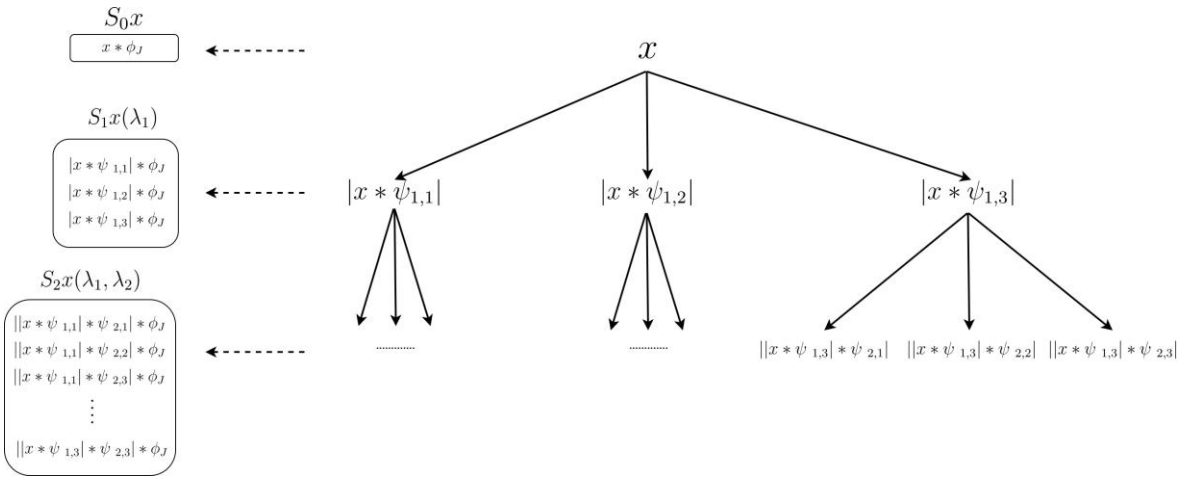
$$S_1x(\lambda_1) = \left|x * \psi_{\lambda_1}\right| * \phi_J \tag{2.19}$$

$N \times N$ boyutundaki bir gri seviye giriş görüntüsü için birinci derece saçılım çıktısı S_1x, her biri $N/2^J \times N/2^J$ uzamsal çözünürlüğe sahip $J \times L$ adet öznitelik haritasından oluşmaktadır. İkinci derece saçılım katsayıları, birinci derece saçılım çıktılarının modülüs çıktısına ilave bir dalgacık dönüşümü uygulanmasıyla elde edilerek, hiyerarşik yapı daha ileri bir aşamaya genişletilmektedir.

$$S_2x(\lambda_1, \lambda_2) = \left|\left|x * \psi_{\lambda_1}\right| * \psi_{\lambda_2}\right| * \phi_J \tag{2.20}$$

$N \times N$ boyutundaki bir gri seviye giriş görüntüsü için ikinci derece saçılım çıktısı S_2x, her biri $N/2^J \times N/2^J$ uzamsal çözünürlüğe sahip $\frac{J(J-1)}{2} \cdot L^2$ adet öznitelik haritasından oluşmaktadır. Nihai saçılım temsili; çoklu ölçek ve yönelimlerdeki yerel doku ve yapıları yakalayan S_0x, S_1x ve S_2x saçılım katsayılarının birleştirilmesi (concatenation) ile elde

edilmektedir. $J = 2$ ölçek ve $L = 3$ yönelime sahip bir Dalgacık Saçılım Ağı örneği Şekil 2.40'ta gösterilmiştir.



Şekil 2.40: Dalgacık Saçılım Ağı genel yapısı.

2.7 Parametrik Saçılım Ağları

Klasik dalgacık saçılım dönüşümlerinde, kullanılan dalgacık filtreleri, parametrik bir anne dalgacıktan türetilmekte ve enerji korunumu gibi teorik özellikleri garanti eden sıkı çerçeve (tight frame) koşulunu sağlayacak biçimde analitik olarak önceden tanımlanmaktadır. Bu sıkı çerçeve koşulu genellikle dalgacık parametrelerinin önceden tanımlanmış analitik bir şemaya göre seçilmesiyle sağlanmaktadır.

Bu yaklaşım, saçılım temsillerine güçlü kuramsal garantiler kazandırsa da, filtre bankasının veriye veya göreve özgü olarak uyarlanamaması önemli bir sınırlama oluşturmaktadır. Bu motivasyondan hareketle geliştirilen Parametrik Saçılım Ağları (PSA) [65], klasik scattering mimarisini korurken, dalgacık filtrelerinin belirli parametrelerini öğrenilebilir hale getirerek bu kısıtı aşmayı amaçlamaktadır. PSA yaklaşımı, özellikle Morlet dalgacıklarını temel almakta ve bu dalgacıkların yapısal parametrelerinin eğitim süreci boyunca geri yayılım ile optimize edilmesine olanak tanımaktadır. Bu yaklaşım, Morlet dalgacıkları temel alınarak, Gauss pencere ölçeği σ, yönelim θ, frekans ölçeği ξ ve en–boy oranı γ parametrelerinin eğitim sürecinde öğrenilmesini mümkün kılmakta ve böylece saçılım dönüşümünün göreve özgü olarak uyarlanmasını sağlamaktadır. Burada R_θ dalgacığın uzaysal yönelimini belirleyen dönme matrisi, D_γ anizotropik ölçeklendirmeyi kontrol eden diyagonal matris ve β sıfır ortalama koşulunu sağlayan bir normalizasyon sabiti olmak üzere, Morlet dalgacıklarının matematiksel ifadesi, Denklem 2.21'de verilmiştir:

$$\psi_{\sigma,\theta,\xi,\gamma}(u) = \exp\left(-\frac{|D_\gamma R_\theta u|^2}{2\sigma^2}\right)\left(e^{i\xi u'} - \beta\right) \tag{2.21}$$

Optimizasyon süreci için kararlı bir başlangıç noktası sağlamak amacıyla, PSA dalgacık parametrelerini bir sıkı çerçeve şemasına dayalı olarak başlatmaktadır. Bu başlangıçtan itibaren parametreler, geri yayılım yoluyla uçtan uca optimize edilmekte ve böylece saçılım temsillerinin verinin yapısına uyum sağlaması mümkün hale gelmektedir.

3. İLİŞKİLİ ÇALIŞMALAR

3.1 Görüntü Bölütleme için Öz-Denetimli Öğrenme Yaklaşımları

Bu bölümde, görüntü bölütleme problemleri için öz-denetimli öğrenme yaklaşımları ayrıntılı biçimde ele alınmaktadır. Bölümün amacı, denetimli ön-eğitimin sınırlılıklarını ortaya koymak, öz-denetimli öğrenmenin bu alandaki motivasyonunu açıklamak ve literatürde önerilen başlıca yöntemleri sistematik bir çerçevede sunmaktır. Anlatım, hem genel bilgisayarlı görü literatürünü hem de tıbbi görüntüleme odaklı çalışmaları kapsayacak şekilde yapılandırılmıştır.

3.1.1 Denetimli ön-eğitimin sınırlılıkları

Derin öğrenme tabanlı anlamsal görüntü bölütleme yöntemlerinde yaygın yaklaşım, kodlayıcı ağların ImageNet gibi büyük ölçekli, etiketli doğal görüntü veri kümeleri üzerinde ön-eğitilmesi ve ardından hedef bölütleme görevine ince ayar (fine-tuning) ile uyarlanmasıdır. FCN [22], SegNet [23], U-Net [24] ve DeepLab [66] gibi mimarilerde bu stratejinin performansı artırdığı gösterilmiştir. Ancak bu yaklaşım, iki temel sınırlamayı beraberinde getirmektedir.

Birincisi, ImageNet benzeri veri kümeleri ile hedef alan (özellikle tıbbi görüntüler) arasındaki belirgin alan uyumsuzluğudur [3]. Doğal görüntüler renk, doku ve nesne çeşitliliği açısından zengin iken; tıbbi görüntüler çoğunlukla tek kanallı, düşük kontrastlı ve anatomik yapılara özgü ince ayrıntılar içermektedir. Bu durum, denetimli ön-eğitimle öğrenilen temsillerin hedef görev için her zaman optimal olmamasına yol açmaktadır. İkinci olarak, denetimli ön-eğitim stratejisi, büyük ölçekli ve yüksek maliyetli etiketli veri kümelerine bağımlıdır [2,4]. Bu bağımlılık, etiketlemenin zor ve pahalı olduğu tıbbi görüntüleme gibi alanlarda ciddi bir kısıt oluşturmaktadır.

1.1.1 Görüntü bölütleme için öz-denetimli öğrenmenin motivasyonu

Öz-denetimli öğrenme, manuel etiketlere ihtiyaç duymadan verinin kendi iç yapısından gözetim sinyalleri üreterek temsil öğrenmeyi amaçlayan bir yaklaşımdır [67]. Bu yaklaşımda, model bir yardımcı görev üzerinden eğitilmekte ve öğrenilen temsiller daha sonra asıl görev olan görüntü bölütleme için kullanılmaktadır. Öz-denetimli öğrenmenin temel motivasyonu, büyük miktarda etiketlenmemiş veriden faydalanarak, alan-özgü ve genellenebilir temsiller öğrenebilmektir.

Görüntü bölütleme bağlamında öz-denetimli öğrenme, özellikle piksel seviyesinde etiket gereksinimini azaltması nedeniyle dikkat çekmektedir. Literatürde yapılan çalışmalar, öz-denetimli ön-eğitimin yalnızca sınıflandırma değil, aynı zamanda yoğun tahmin (dense prediction) görevlerinde de önemli kazanımlar sağlayabileceğini göstermektedir [68].

3.1.2 Görüntü seviyesinde öz-denetimli öğrenme yaklaşımları

Görüntü seviyesinde öz-denetimli öğrenme yaklaşımları, genellikle global temsil öğrenimine odaklanmaktadır. Bu yöntemlerde kodlayıcı ağ, karşılaştırmalı veya karşılaştırmalı olmayan SimCLR, MoCo, BYOL ve Barlow Twins gibi öz denetimli öğrenme yöntemleriyle eğitilmekte ve öğrenilen temsiller daha sonra bölütleme ağının kodlayıcı kısmını başlatmak için kullanılmaktadır.

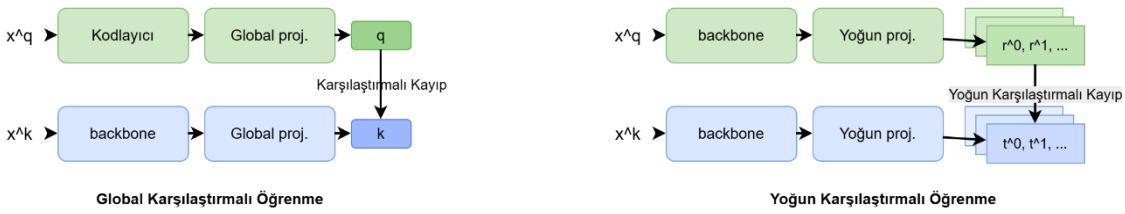
Bu yaklaşımda yaygın strateji, bölütleme mimarisinin yalnızca kodlayıcı kısmının öz-denetimli olarak ön-eğitilmesi, kod çözücü kısmının ise rastgele başlatılarak denetimli olarak eğitilmesidir. Literatürde yapılan deneyler, özellikle sınırlı etiketli veri senaryolarında bu yaklaşımın rastgele başlatmaya kıyasla belirgin performans artışları sağladığını göstermektedir. Bu yaklaşıma örnek olarak, Zeng ve arkadaşları [69], hacimsel veri kümelerinde ardışık görüntü dilimleri arasındaki doğal konumsal ilişkilerden faydalanarak, uzamsal bağlamı yakalamayı hedefleyen karşılaştırmalı bir öğrenme çerçevesi önermiştir. BT-UNet [70] çalışmasında ise Barlow Twins tabanlı bir öz-denetimli öğrenme yaklaşımı, U-Net mimarisinin kodlayıcı kısmı ile birleştirilmiş ve etiketlenmemiş biyomedikal görüntülerden fazlalığı azaltılmış ve kararlı temsiller öğrenilebileceği gösterilmiştir. Daha yakın tarihli bir çalışmada, Kalapos ve Gyires-Tóth [71], BYOL ile ön-eğitilmiş bir kodlayıcının U-Net mimarisinde kullanılmasının tıbbi görüntü bölütleme performansını anlamlı ölçüde iyileştirdiğini raporlamıştır. Bu çalışmalar, görüntü seviyesinde öz-denetimli ön-eğitimin, özellikle sınırlı etiketli veri senaryolarında, bölütleme ağları için etkili bir başlangıç noktası sunduğunu ortaya koymaktadır. Bu tez çalışmasında da benzer bir öz-denetimli ön-eğitim stratejisi benimsenmiş; ancak mevcut yaklaşımlardan farklı olarak, BYOL tabanlı kodlayıcının erken aşamaları matematiksel olarak temellendirilmiş saçılım tabanlı bir ön-uç ile değiştirilerek, daha yapısal ve kararlı özniteliklerin öğrenilmesi hedeflenmiştir.

3.1.3 Piksel seviyesinde öz-denetimli öğrenme yaklaşımları

Görüntü bölütleme, her bir pikselin ait olduğu sınıfın belirlenmesini gerektiren yoğun bir tahmin problemi olduğundan, öğrenilen temsillerin yalnızca grobal görüntü

istatistiklerini değil, aynı zamanda yerel uzamsal yapı ve sınır bilgilerini de etkin biçimde kodlaması beklenmektedir. Bu bağlamda, görüntü seviyesinde öz-denetimli öğrenme yaklaşımlarının çoğunlukla global görüntü temsillerine odaklanması, ince yapısal ayrıntıların ve nesne sınırlarının modellenmesinde sınırlı kalabilmektedir. Görüntü seviyesinde öz-denetimli öğrenme yaklaşımlarından farklı olarak, piksel seviyesinde öz-denetimli öğrenme yöntemleri, öz-denetimli kayıp fonksiyonlarını doğrudan yerel öznitelikler veya pikseller üzerinde tanımlayarak uzamsal olarak tutarlı ve anlamsal açıdan zengin temsilleri öğrenmeyi hedeflemektedir. Bu yaklaşımlar, özellikle segmentasyon gibi yoğun tahmin görevlerinde gerekli olan ince taneli (fine-grained) anlamsal ve yapısal bilgilerin öğrenilmesini teşvik etmektedir.

Dense Contrastive Learning (DenseCL) [72] bu yaklaşımın öne çıkan örneklerinden biridir. DenseCL, örnek ayrımı (instance discrimination) yaklaşımını yoğun tahmin görevlerine genişleterek, aynı görüntünün farklı veri artırımları altında elde edilen görünümlerindeki piksel düzeyindeki öznitelikleri birbiriyle eşleştirmeyi; farklı görüntülere ait pikselleri ise karşıt örnekler olarak ele almayı hedefleyen bir öz-denetimli öğrenme yaklaşımı önermektedir. Şekil 3.1'de DenseCL'nin genel mimarisi gösterilmiştir. Benzer şekilde PixPro [73], pikselden-piksele tutarlılık ilkesine dayalı bir kayıp fonksiyonu tanımlayarak, yerel yapısal bilgilerin öğrenilmesini sağlayan bir yömtem önermektedir. Bir diğer çalışmada ise Chaitanya ve arkadaşları [74], hem global görüntü temsillerini hem de yerel piksel-seviyesindeki öznitelikleri birlikte kullanan bir karşılaştırmalı öğrenme yaklaşımı önererek, özellikle doğru bölütleme için kritik öneme sahip alan-özgü yapısal örüntülerin öğrenilmesini mümkün kıldığını göstermiştir.



Şekil 3.1: DenseCL yönteminin genel yapısı.

3.2 Saçılım Tabanlı Hibrit Derin Öğrenme Modelleri

DSA, ötelemeye karşı değişmezlik ve küçük deformasyonlara karşı kararlılık özellikleri sayesinde, el yazısı tanıma ve doku analizi gibi düşük seviyeli görsel içerik analizlerinde başarılı sonuçlar elde etmiştir. Buna karşın, daha karmaşık yapısal örüntülerin ve yüksek seviyeli anlamsal bilgilerin söz konusu olduğu problemlerde, sabit filtre yapısının

temsil gücünü sınırladığı ve modelin genellenebilirliğini kısıtladığı görülmektedir [75]. Bu kısıtlamaları aşmak amacıyla, DSA'nın yapısını eğitilebilir ESA katmanlarıyla birleştiren hibrit mimariler yaygın olarak araştırılmaya başlanmıştır. Bu hibrit yaklaşımlarda, DSA genellikle sabit bir ön-yüz öznitelik çıkarıcı olarak konumlandırılmakta ve ardından öğrenilebilir evrişimsel katmanlar aracılığıyla temsilin göreve uyumlu biçimde uyarlanması sağlanmaktadır. Oyallon ve arkadaşları [76,77], bu tür DSA–ESA hibrit mimarilerinin, daha az öğrenilebilir parametreye sahip olmalarına rağmen rekabetçi doğruluk seviyelerine ulaşabildiğini ve genelleme performansında iyileşme sağladığını göstermiştir.

Saçılım dönüşümünün esnekliğini ve veriye adaptasyon kapasitesini artırmak adına daha ileri düzey yöntemler de geliştirilmiştir. Örneğin, Cotter ve Kingsbury [78], sabit saçılım katsayılarının öğrenilebilir ağırlıklarla modüle edildiği yerel değişmez evrişimli katmanlar önererek dönüşüme adaptasyon yeteneği kazandırmıştır

Denetimli ve hibrit mimarilere ek olarak, dalgacık saçılım dönüşümleri öz-denetimli öğrenme bağlamında da incelenmiştir. Bu doğrultuda önerilen ScatSimCLR [79], DSA bileşenlerini bir öz-denetimli öğrenme çerçevesine entegre eden ilk çalışmalardan biri olarak öne çıkmaktadır. ScatSimCLR yaklaşımı, DSA tabanlı temsilleri karşılaştırmalı SimCLR kayıp fonksiyonu ile birleştirerek, etiketlenmemiş veriler üzerinden anlamlı özniteliklerin öğrenilebileceğini göstermiştir. Bu yönüyle çalışma, saçılım tabanlı temsillerin öz-denetimli öğrenme kapsamında kullanılabilirliğini ortaya koyan önemli bir adım olarak değerlendirilmektedir.

Bununla birlikte, ScatSimCLR ile bu tez kapsamında önerilen yöntem arasında bazı temel farklılıklar bulunmaktadır. Öncelikle, ScatSimCLR karşılaştırmalı öğrenmeye dayalı SimCLR yöntemini kullanırken, bu çalışmada karşılaştırmalı olmayan bir öz-denetimli öğrenme yaklaşımı olan BYOL kullanılmıştır. Ayrıca ScatSimCLR'de saçılım temsilleri tüm kodlayıcı mimarinin yerini alacak şekilde kullanılırken, bu tez çalışmasında yalnızca ResNet-50 kodlayıcısının başlangıç aşamaları saçılım tabanlı bir ön-uç ile değiştirilmiş; böylece daha derin katmanlarda öğrenilebilir evrişimsel temsillerin korunması amaçlanmıştır. Bir diğer önemli fark, öğrenilen temsillerin değerlendirilme biçimidir. ScatSimCLR yaklaşımı temsilleri görüntü sınıflandırma görevi üzerinden değerlendirirken, bu tez çalışmasında öz-denetimli olarak öğrenilen temsiller görüntü bölütleme problemi kapsamında incelenmiştir. Son olarak, ScatSimCLR sabit DSA kullanımıyla sınırlıyken, bu tez çalışması sabit saçılım dönüşümlerinin ötesine geçerek, öğrenilebilir PSA da öz-denetimli öğrenme çerçevesine dahil etmektedir.

Bu çalışmalar, saçılım tabanlı temsillerin güçlü teorik özelliklerinin, öğrenilebilir katmanlarla birleştirildiğinde hem temsil kararlılığını hem de ifade gücünü artırabildiğini ortaya koymaktadır. Bu bağlamda, DSA'nın hibrit mimarilerde kullanımı, sabit ve kararlı bir başlangıç temsili ile temsil öğrenimine dayalı yaklaşımların esnekliğini bir araya getiren etkili bir yaklaşım olarak değerlendirilmektedir.

4. SAÇILIM-TABANLI ÖZ-DENETİMLİ ÖĞRENME İLE ETİKET-VERİMLİ TIBBİ GÖRÜNTÜ BÖLÜTLEME

Derin öğrenme tabanlı modeller, bilgisayarlı görü görevlerinde son yıllarda dikkate değer başarılar göstermiştir; ancak bu başarının temelini oluşturan denetimli öğrenme yaklaşımları, yüksek miktarda manuel olarak etiketlenmiş veri gereksinimi nedeniyle önemli bir sınırlama barındırmaktadır. Görüntü bölütleme özelinde ise her bir pikselin ayrı ayrı etiketlenmesi gerekliliği, bu süreci oldukça zahmetli, zaman alıcı ve maliyetli bir hale getirmektedir. Tıbbi görüntüleme alanında bu zorluk çok daha belirgin hale gelmektedir; görüntülerin elde edilmesinin maliyetli ve zahmetli olması ile doğru etiketleme sürecinin uzman klinik bilgi gerektirmesi, yüksek kaliteli etiketli veri üretimini önemli ölçüde zorlaştırmaktadır.

Literatürde bu sorunun çözümünde yaygın olarak başvurulan ImageNet tabanlı transfer öğrenme yöntemleri, doğal görüntüler ile tıbbi görüntüler arasındaki belirgin alan farkı (domain gap) nedeniyle her zaman beklenen başarıyı sergileyememektedir. Bu durum, yüksek kapasiteli denetimli modellerin sınırlı veri rejimlerinde genelleme performansının düşmesine yol açmakta ve etiket-verimli (label-efficient) öğrenme stratejilerini zorunlu hale getirmektedir.

Bu çalışma kapsamında ele alınan temel problem, sınırlı miktarda etiketli tıbbi görüntü verisiyle yüksek doğrulukta ve kararlı bölütleme modellerinin nasıl geliştirilebileceği sorusu etrafında şekillenmektedir. Bu bağlamda, önceki bölümlerde ayrıntılı olarak incelenen öz-denetimli öğrenme yaklaşımları, etiketlenmemiş verilerden temsil öğrenme potansiyeli sunarak bu probleme güçlü bir çözüm alternatifi sağlamaktadır. Ancak mevcut öz-denetimli öğrenme yaklaşımları yüksek sayıda öğrenilebilir parametreye sahip derin ağlara dayandığından hem hesaplama maliyeti hem de sınırlı veri senaryolarında kararlılık açısından önemli kısıtlar barındırmaktadır.

Belirtilen sınırlamaları aşmak amacıyla, öz-denetimli öğrenme yaklaşımı ile dalgacık saçılım temelli mimarileri birleştiren bir yaklaşım önerilmektedir. Bu bağlamda Öz-denetimli öğrenme kodlayıcısının ilk aşamalarında DSA ve PSA kullanılmıştır. Bu yaklaşım, derin ağların ilk katmanlarında öğrenilen filtreler yerine, öteleme değişmezliği ve küçük deformasyonlara karşı kararlılık gibi matematiksel kanıtlanmış özelliklere sahip saçılım temsillerinin kullanılmasını mümkün kılmaktadır.

Önerilen saçılma-tabanlı ön-uç (front-end) yapısı, bir yandan öğrenilebilir parametre sayısını önemli ölçüde azaltırken, diğer yandan öz-denetimli öğrenme süreci için daha kararlı ve yapısal olarak anlamlı bir başlangıç temsili sunmaktadır. Modelin başlangıç aşamasında sahip olduğu bu güçlü matematiksel ön-bilgiler, sınırlı miktarda etiketli veriyle yapılan eğitimlerde dahi modelin karmaşık dokuları ayırt etmesini kolaylaştırarak etiket-verimli bir öğrenme stratejisini mümkün kılmaktadır. Sabit DSA yapıları güçlü öncül bilgiler sağlarken, PSA mimarisi bu yapıyı sınırlı sayıda öğrenilebilir dalgacık parametresi ile genişleterek veriye özgü uyarlanabilirliği mümkün kılmaktadır. Böylece, tamamen öğrenilebilir derin katmanlar ile tamamen sabit filtre bankaları arasında dengeli bir çözüm elde edilmektedir.

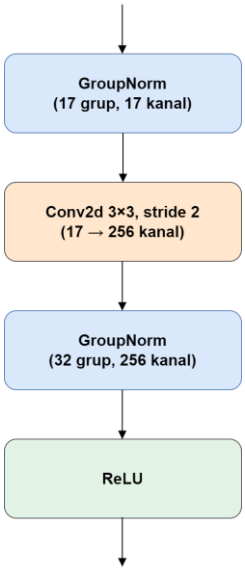
4.1 Önerilen Yöntem

Bu çalışmada önerilen yöntem, saçılım tabanlı öznitelik çıkarıcıların BYOL öz-denetimli öğrenme çerçevesine entegre edildiği ve elde edilen ön-eğitimli kodlayıcının kardiyak görüntü bölütleme görevi için bir U-Net mimarisine aktarıldığı iki aşamalı bir öğrenme yapısından oluşmaktadır. Metodoloji, ilk aşamada etiketlenmemiş veriler üzerinde modifiye edilmiş bir BYOL mimarisi ile gerçekleştirilen öz-denetimli ön-eğitim ve ikinci aşamada öğrenilen kodlayıcının görüntü bölütleme amacıyla U-Net modeli üzerinde denetimli olarak ince ayarlanmasını içermektedir.

Ön eğitim aşamasında, etiketlenmemiş kardiyak görüntülerinden genel amaçlı görüntü temsilleri öğrenmek amacıyla modifiye edilmiş bir BYOL mimarisi kullanılmıştır. Standart BYOL konfigürasyonunda kodlayıcı ağ olarak beş aşamadan oluşan ResNet-50 mimarisi tercih edilmektedir. Ancak bu çalışmada, düşük seviyeli özellik çıkarımının dayanıklılığını artırmak ve matematiksel olarak kanıtlanmış öteleme değişmezliği ve küçük deformasyonlara karşı kararlılık özelliklerini eklemek amacıyla, ResNet-50'nin ilk iki aşaması saçılım tabanlı bir ön-uç ile değiştirilmiştir. Bu mimari değişim kapsamında, geleneksel evrişimli katmanlar yerine iki farklı alternatif uygulanmıştır:

- **DSA Ön-ucu**: Önceden tanımlı Morlet dalgacıkları kullanan, tamamen sabit ve öğrenilebilir parametre içermeyen Dalgacık Saçılım Ağları.
- **PSA Ön-ucu**: Saçılım dönüşümünün temel yapısını korurken, sınırlı sayıda dalgacık parametresinin eğitim süreci boyunca öğrenilebilir hale getirildiği Parametrik Saçılım Ağları.

Saçılım temsillerinin ürettiği öznitelik haritalarının boyutları, ResNet-50 mimarisinin sonraki katmanlarının beklediği giriş boyutlarıyla doğrudan uyumlu olmadığından, bu uyumsuzluğu gidermek amacıyla Öznitelik Hizalama Modülü (ÖHM) kullanılmıştır. ÖHM, saçılım ön-uç ile derin kodlayıcının geri kalan katmanları arasında konumlandırılarak, saçılım özniteliklerinin hem uzamsal çözünürlük hem de kanal boyutu açısından ResNet-50 ile uyumlu hale getirilmesini sağlamaktadır. ÖHM'nin ilk aşamasında, grup sayısı saçılım kanal sayısına eşit olacak şekilde yapılandırılmış bir Group Normalization (GN) işlemi uygulanmıştır. Bu yapılandırma, her bir saçılım katsayısının kanal bazında bağımsız olarak normalize edilmesini sağlamakta; böylece farklı ölçek ve yönelimlere karşılık gelen saçılım kanalları arasında istenmeyen kanal-arası bağımlılıkların oluşması engellenmektedir. Bu normalizasyon stratejisi, saçılım temsillerinin yapısıyla uyumludur; çünkü her bir kanal farklı bir ölçek ve yönelime karşılık gelen bağımsız bir yanıtı temsil etmektedir. Normalizasyonun ardından, filtre boyutu 3×3, adım uzunluğu 2 ve doldurma değeri 1 olan iki boyutlu bir evrişimsel projeksiyon katmanı uygulanmaktadır. Bu katman, bir yandan adım uzunluğu işlemi sayesinde öznitelik haritalarının uzamsal çözünürlüğünü ResNet-50'nin Aşama 3 katmanının giriş gereksinimleriyle uyumlu hale getirirken, diğer yandan 3×3 evrişim işlemi aracılığıyla saçılım katsayılarını, sonraki artık bloklar tarafından beklenen 256 kanallı öznitelik uzayına yansıtan bir projeksiyon işlevi görmektedir. Projeksiyon katmanını takiben, 32 grup içeren ikinci bir Group Normalization katmanı uygulanmakta ve ardından ReLU aktivasyon fonksiyonu kullanılmaktadır. Bu yapılandırma, derin artık ağlarda yaygın olarak kullanılan normalizasyon uygulamalarıyla uyumlu olup, öz-denetimli tıbbi görüntüleme senaryolarında sıklıkla karşılaşılan küçük yığın boyutları altında kararlı bir eğitim süreci sağlamaktadır. Bu tasarım sayesinde model, ilk katmanlarda saçılım dönüşümünün sağladığı öteleme değişmez ve küçük deformasyonlara karşı kararlı düşük seviyeli temsillerden yararlanırken, derin katmanlarda (Aşama 3, 4 ve 5) yüksek seviyeli anlamsal bilgileri öğrenme yeteneğini korumaktadır. Öznitelik Hizalama Modülünün mimarisi Şekil 4.1'de görselleştirilmiştir.



Şekil 4.1: Öznitelik Hizalama Modülü genel yapısı.

Ayrıca bu mimari değişim, Öznitelik Hizalama Modülü dahil edildiğinde dahi modelin toplam öğrenilebilir parametre sayısında da anlamlı bir azalma sağlamaktadır. Standart BYOL mimarisinde ResNet-50 mimarisinin ilk iki aşaması toplamda 219.072 öğrenilebilir parametre içermektedir. Önerilen BYOL+DSA ve BYOL+PSA mimarilerinde bu aşamalar tamamen kaldırılarak, yerlerine sırasıyla öğrenilebilir parametre içermeyen bir Dalgacık Saçılım Ağı ve yalnızca 68 öğrenilebilir parametre içeren Parametrik Saçılım Ağı yerleştirilmiştir. Saçılım tabanlı ön-uç ile ResNet-50'nin üçüncü aşaması arasında konumlandırılan Öznitelik Hizalama Modülü ise 39.714 öğrenilebilir parametre içermektedir. Bu yapılandırma sonucunda, önerilen mimariler standart BYOL-ResNet-50 mimarisine kıyasla yaklaşık 180K daha az öğrenilebilir parametre ile çalışmaktadır. İlgili mimariler arasındaki parametre sayısı karşılaştırması Çizelge 4.1'de sunulmaktadır.

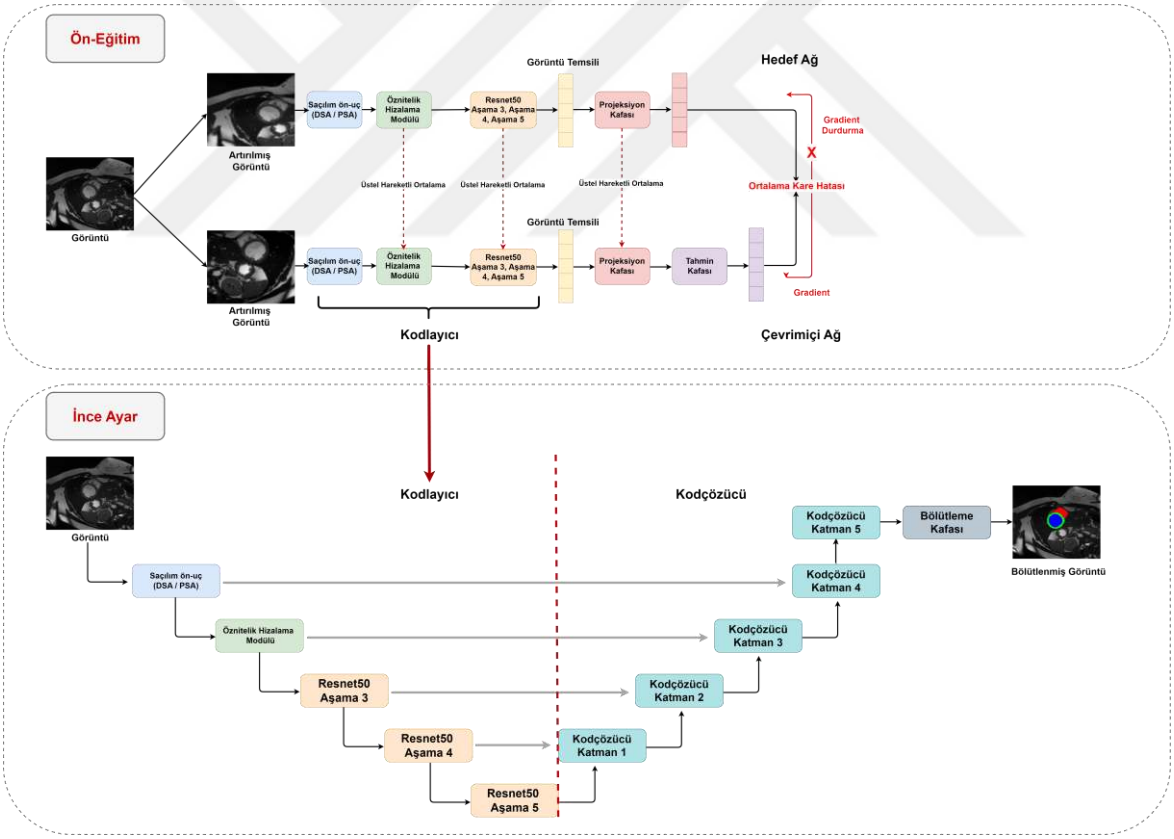
Çizelge 4.1: Önerilen mimarilerin standart BYOL mimarisi ile parametre sayısı karşılaştırması.

Mimari	Çıkarılan (Aşama 1+2)	Eklenen (ÖHM)	Eklenen Öğrenebilir Parametre Sayısı	Parametre Sayısındaki Net Azalma
BYOL+DSA	219.072	39.714	0	179.358
BYOL+PSA	219.072	39.714	68	179.290

Öz-denetimli öğrenme ile eğitilen kodlayıcı, daha sonra kardiyak görüntü bölütleme görevi için doğrudan bir U-Net mimarisine aktarılmaktadır. Ön-eğitim sırasında kullanılan kodlayıcı yapısı, saçılım tabanlı ön-uç ve ResNet'in kalan aşamalarıyla birlikte değiştirilmeden korunmakta, böylece ağırlıkların doğrudan aktarımı mümkün olmaktadır.

Bu transfer sürecinde, U-Net'in kod çözücü kısmı standart tasarımını korumakta ve ön eğitim sürecine dahil olmadığı için rastgele değerlerle başlatılmaktadır.

Bölütleme ağı, giriş olarak iki boyutlu gri-seviye kardiyak görüntüleri almakta ve çok sınıflı piksel seviyesinde bölütleme maskeleri üretmektedir. Eğitim sürecinde, Dice kaybı ile çapraz entropi kaybının birleşiminden oluşan bir kayıp fonksiyonu kullanılmaktadır. Çapraz entropi kaybı piksel bazlı sınıflandırma doğruluğunu teşvik ederken, Dice kaybı tahmin edilen ve gerçek maske arasındaki bölgesel örtüşmeyi doğrudan maksimize etmektedir. Bu iki kaybın birlikte kullanılması, hem sınıf dengesizliği problemini hafifletmekte hem de sınır bölgelerinde daha tutarlı tahminler elde edilmesini sağlamaktadır. Sonuç olarak, dalgacık saçılımının teorik kararlılığı ile derin öğrenmenin esnek temsil yeteneğini birleştiren bu yöntem, özellikle sınırlı etiketli verinin bulunduğu durumlarda yüksek doğruluğa sahip ve etiket-verimli bölütleme sonuçları üretmektedir. Şekil 4.2'de önerilen yöntem gösterilmiştir.



Şekil 4.2: Önerilen yöntem.

4.2 Deneysel Çalışmalar

Bu bölümde, önerilen saçılım tabanlı öz-denetimli öğrenme yönteminin etkinliğini değerlendirmek amacıyla yürütülen deneysel çalışmaların detayları sunulmaktadır. Çalışma

kapsamında kullanılan veri seti, uygulama aşamasında tercih edilen deneysel kurulum ve değerlendirme metrikleri ayrıntılı olarak ele alınmıştır. Tüm deneyler, önerilen yöntemin etiket-verimliliğini ve genelleme yeteneğini kapsamlı bir şekilde analiz edecek biçimde tasarlanmıştır.

4.2.1 Veri seti

Bu çalışma kapsamında gerçekleştirilen deneylerde, 2017 Automated Cardiac Diagnosis Challenge (ACDC) [80] kapsamında sunulan, açık erişimli kardiyak cine-MR görüntü veri seti kullanılmıştır. Veri seti, kısa eksen (short-axis) düzleminde elde edilmiş cine-MR görüntülerinden oluşan toplam 150 hasta verisini içermektedir. Resmi yarışma konfigürasyonuna uygun olarak, bu hastaların 100'ü eğitim/doğrulama, 50'si ise ayrılmış test (held-out test) kümesi olarak belirlenmiştir. Veri setindeki hastalar, normal, miyokard enfarktüsü, dilate kardiyomiyopati, hipertrofik kardiyomiyopati ve anormal sağ ventrikül olmak üzere beş farklı klinik alt gruba eşit olarak dağılmıştır. Görüntüler, 1.5T ve 3.0T MR cihazları kullanılarak yaklaşık altı yıllık bir zaman diliminde elde edilmiş olup, uzamsal çözünürlükleri piksel başına 1.22–1.68 mm² aralığında değişmektedir. Her bir hasta için, kalp döngüsünün tamamını kapsayan çok kesitli 3 boyutlu cine-MR hacimleri mevcuttur. Sol ventrikül (LV), sağ ventrikül (RV) ve miyokard (MYO) için bölütleme etiketleri yalnızca end-diyastolik (ED) ve end-sistolik (ES) fazlarda sağlanmıştır. Kardiyak döngünün geri kalan fazları etiketlenmemiş olup, bu görüntüler öz-denetimli ön-eğitim aşamasında kullanılmıştır. Veri seti, kesit bazında önemli yapısal çeşitlilik de içermekte; bazı kesitlerde kalp yapıları tamamen bulunmazken, bazı kesitlerde farklı yapı kombinasyonları yer almaktadır. Bu özellik, veri setini öz-denetimli öğrenme yaklaşımlarının değerlendirilmesi açısından özellikle uygun hale getirmektedir.

Bu çalışma kapsamında ACDC veri setine ek olarak, modellerin farklı görüntüleme modaliteleri ve daha yüksek anatomik değişkenlik altında değerlendirilmesi amacıyla açık erişimli Congenital Heart Disease (CHD) veri seti de kullanılmıştır [81]. CHD veri seti, konjenital kalp hastalığı tanısı almış hastalardan elde edilen kardiyak bilgisayarlı tomografi (BT) görüntülerinden oluşmaktadır. Veri seti başlangıçta 68 adet üç boyutlu kardiyak BT hacmi içermekte olup, uzamsal çözünürlük tutarsızlığı nedeniyle bir vaka çalışma dışı bırakılmış ve deneylerde toplam 67 hasta verisi kullanılmıştır. CHD veri seti, 14 farklı konjenital kalp hastalığı tipini kapsamakta olup, ACDC veri setine kıyasla belirgin derecede daha yüksek anatomik çeşitlilik ve yapısal karmaşıklık sunmaktadır. Her bir hasta için üç

boyutlu kardiyak BT hacimleri üzerinde, sol ventrikül (LV), sağ ventrikül (RV), sol atriyum (LA), sağ atriyum (RA), miyokard (MYO), aorta ve pulmoner arter olmak üzere toplam yedi farklı kardiyak yapı için manuel, voksel düzeyinde bölütleme etiketleri sağlanmıştır. Bu özellikleriyle CHD veri seti, özellikle karmaşık anatomilere sahip hasta gruplarında bölütleme performansının değerlendirilmesi açısından zorlu ve tamamlayıcı bir test ortamı sunmaktadır.

4.2.2 Deneysel kurulum

Tüm deneyler PyTorch derin öğrenme çatısı kullanılarak gerçekleştirilmiş ve hesaplamalar NVIDIA A100 GPU üzerinde yürütülmüştür. Deneysel süreç, öz-denetimli ön-eğitim ve denetimli bölütleme için ince ayar olmak üzere iki ana aşamadan oluşmaktadır.

4.2.2.1 Öz-denetimli ön-eğitim

Öz-denetimli temsil öğrenme aşamasında, karşılaştırmalı olmayan bir öz-denetimli öğrenme yöntemi olan Bootstrap Your Own Latent (BYOL) mimarisi kullanılmıştır. Ön-eğitim süreci, hem ACDC hem de CHD veri setlerine ait etiketlenmemiş görüntü kesitleri kullanılarak gerçekleştirilmiştir.

ACDC veri seti için, eğitim kümesinde yer alan 100 hastaya ait dört boyutlu (4B) cine-MR hacimlerinden elde edilen tüm görüntü kesitleri, herhangi bir bölütleme etiketi kullanılmaksızın ön-eğitim aşamasına dahil edilmiştir. Bu kapsamda, kardiyak döngünün tüm fazlarına ait kesitler kullanılmış; yalnızca end-diyastolik (ED) ve end-sistolik (ES) fazlarda mevcut olan etiketler bu aşamada bilinçli olarak göz ardı edilmiştir. Sonuç olarak, ACDC veri seti üzerinde gerçekleştirilen ön-eğitim sürecinde toplam 27.253 adet etiketlenmemiş MR kesiti kullanılmıştır.

CHD veri seti için ise, eğitim kümesinde yer alan 50 hastaya ait üç boyutlu kardiyak BT hacimlerinden elde edilen tüm görüntü kesitleri, benzer şekilde herhangi bir bölütleme bilgisine erişilmeksizin ön-eğitim sürecine dahil edilmiştir. Bu işlem sonucunda, CHD veri setinden toplam 13.188 adet etiketlenmemiş görüntü kesiti öz-denetimli ön-eğitim aşamasında kullanılmıştır.

Bu aşamada üç farklı kodlayıcı mimarisi karşılaştırılmıştır:

1. Standart ResNet-50 tabanlı BYOL kodlayıcısı
2. Aşama 1 ve Aşama 2'nin DSA ile değiştirildiği ResNet-50

3. Aşama 1 ve Aşama 2'nin PSA ile değiştirildiği ResNet-50

DSA yapısı, Kymatio kütüphanesi kullanılarak uygulanmış ve kardiyak görüntülerin görüntülerinin ince yapısal detaylarını korumak amacıyla tek ölçekli (J = 1) ve çok yönelimli (L = 16) bir yapı tercih edilmiştir. Bu yapı toplam 17 adet saçılım katsayısı üretmekte olup, bu katsayılar ÖHM aracılığıyla 256 kanala projekte edilerek ResNet-50'nin üçüncü aşamasın uyumlu hale getirilmiştir. PSA tabanlı yapı ise aynı mimari düzeni korurken, dalgacık parametrelerinin eğitim süreci boyunca öğrenilebilir olmasına imkan tanımaktadır.

BYOL çerçevesinde kullanılan projeksiyon ve tahmin kafaları, gizli katman boyutu 4096 olan çok katmanlı algılayıcılardan oluşmaktadır. Optimizasyon için LARS algoritması kullanılmış; öğrenme oranı 400 epoch boyunca kosinüs azalma (cosine decay) ile planlanmış ve ilk 10 epoch boyunca doğrusal ısınma (warm-up) uygulanmıştır. Eğitimler 128 yığın boyutu gerçekleştirilmiş ve doğrusal ölçekleme kuralına uygun olarak temel öğrenme oranı 0.3 olarak belirlenmiştir. Hedef kodlayıcı için kullanılan EMA momentumu 0.99'dan başlatılmış ve eğitim boyunca 1.0'a doğru kosinüs planı ile artırılmıştır.

Veri artırma işlemleri; rastgele kırpma ve yeniden boyutlandırma, rastgele yatay çevirme, renk bozulmaları ve Gauss bulanıklaştırma adımlarını içermektedir. Kardiyak yapıların mekansal bütünlüğünü korumak amacıyla kırpma işlemi, orijinal BYOL ayarlarına kıyasla daha dar bir ölçek aralığında gerçekleştirilmiştir. Tüm görüntüler, ön-eğitim öncesinde sabit bir uzamsal çözünürlüğe ($224\ x\ 224$) yeniden boyutlandırılmıştır.

4.2.2.2 Görüntü bölütleme için ince ayar

Öz-denetimli ön-eğitim tamamlandıktan sonra, öğrenilen kodlayıcı ağırlıkları U-Net mimarisi içerisine aktarılmış ve denetimli bölütleme görevinde kullanılmıştır. ACDC veri seti üzerinde gerçekleştirilen denetimli bölütleme deneylerinde, yalnızca etiketli olan end-diyastolik (ED) ve end-sistolik (ES) fazlara ait görüntü kesitleri kullanılmıştır. Eğitim kümesinde yer alan 100 hasta için bu durum, hasta başına ortalama yaklaşık 19 kesit olmak üzere toplam 1.902 adet etiketli MR kesitine karşılık gelmektedir.

Değerlendirme sürecinde, patoloji dağılımını koruyan ve hasta bazında gerçekleştirilen katmanlı 5-katlı çapraz doğrulama stratejisi benimsenmiştir. Her bir çapraz doğrulama katında, ACDC veri setinde yer alan beş farklı klinik patoloji grubunun dağılımı korunacak şekilde eğitim ve doğrulama kümeleri oluşturulmuştur. Çapraz doğrulama süreci, üç farklı rastgele başlangıç (random seed) ile tekrarlanmış; her tekrarda yeni eğitim/doğrulama bölmeleri üretilerek her yöntem için toplam 15 bağımsız eğitim çalışması

gerçekleştirilmiştir. Bu yaklaşım, sonuçların istatistiksel olarak daha güvenilir ve genellenebilir olmasını sağlamaktadır.

Veri sızıntısını (data leakage) önlemek amacıyla tüm veri bölmeleri hasta bazında yapılmış; aynı hastaya ait görüntü kesitlerinin aynı çapraz doğrulama katı içerisinde eşzamanlı olarak eğitim ve doğrulama kümelerinde yer almasına kesinlikle izin verilmemiştir. Her bir kat için, doğrulama kümesi üzerinde en iyi performansı gösteren model seçilmiş ve bu modeller, 50 hastadan oluşan ayrılmış test kümesi üzerinde değerlendirilmiştir. Test kümesi toplam 1.076 adet etiketli MR kesiti içermekte olup, nihai performans sonuçları ortalama Dice ve Jaccard katsayıları (ortalama ± standart sapma) biçiminde raporlanmıştır.

Önerilen modellerin etiket verimliliğini değerlendirebilmek amacıyla, önceki çalışmalarda da kullanılan etiket-kısıtlı ince ayar stratejisi benimsenmiştir. Bu kapsamda, her bir çapraz doğrulama katında, eğitim kümesinde yer alan hastalar arasından rastgele olarak M adet hasta seçilmiş ve ince ayar sürecinde yalnızca bu hastalara ait bölütleme maskeleri kullanılmıştır.

Ön-eğitim aşamasının, etiketli verinin sınırlı olduğu bölütleme senaryolarındaki etkisini kapsamlı biçimde inceleyebilmek amacıyla, $M \in \{5, 10, 20, 40, 80\}$ olmak üzere farklı etiketli hasta sayıları ile deneyler gerçekleştirilmiştir [69,82,83]. Her bir M değeri için oluşturulan etiketli alt küme, ilgili çapraz doğrulama bölmesinin eğitim kısmından seçilmiş; örnekleme sürecinde, patoloji dağılımının dengeli biçimde korunmasına özen gösterilmiştir.

CHD veri seti için denetimli bölütleme ince ayarı, hasta bazında oluşturulan bir veri bölmesi kullanılarak gerçekleştirilmiştir. Bu kapsamda, 50 hasta eğitim ve doğrulama amacıyla ayrılmış; kalan 17 hasta ise ayrılmış test (held-out test) kümesi olarak kullanılmıştır. Eğitim/doğrulama alt kümesi üzerinde 5-katlı çapraz doğrulama uygulanmış ve tüm veri bölmeleri, kesit düzeyinde veri sızıntısını önlemek amacıyla hasta bazında gerçekleştirilmiştir.

ACDC veri setinden farklı olarak, CHD veri setindeki her bir üç boyutlu kardiyak BT hacmi, tüm kesitler için yoğun voksel düzeyinde bölütleme etiketleri içermektedir. Ancak, veri setleri arasında etiket verimliliğinin adil bir biçimde karşılaştırılabilmesi amacıyla, hasta başına kullanılan denetimli bilgi miktarı sınırlandırılmıştır. Bu doğrultuda, her bir CHD hastası için hacmin aksiyel doğrultusu boyunca sabit aralıklarla uniform örnekleme yapılarak 40 adet etiketli kesit seçilmiş; böylece tüm kardiyak anatomiyi

kapsayan ancak denetimli bilgi miktarı kontrollü bir eğitim senaryosu oluşturulmuştur [84]. Bu yaklaşım sonucunda, tüm 50 eğitim hastası kullanıldığında denetimli ince ayar sürecinde toplam 2.000 adet etiketli görüntü kesiti elde edilmiştir.

Etiket-kısıtlı deneyler, $M \in \{2, 5, 10, 20, 40\}$ olacak şekilde farklı etiketli hasta sayıları ile gerçekleştirilmiş; burada M, eğitim sürecinde kullanılan etiketli hasta sayısını ifade etmektedir. Nihai değerlendirme yalnızca ayrılmış test kümesi üzerinde yapılmış olup, test kümesi toplam 4.088 adet etiketli kesit içeren 17 hastadan oluşmaktadır.

İnce ayar sürecinde, Dice kaybı ve çapraz entropi kaybının birleşiminden oluşan bir kayıp fonksiyonu kullanılmış ve Adam optimizasyon yöntemi ile 10^{-4} başlangıç öğrenme oranı tercih edilmiştir. Öğrenme oranı, 10 epoch'luk ısınma evresinin ardından kosinüs planı ile azaltılmış ve eğitimler toplam 150 epoch boyunca sürdürülmüştür.

4.2.2.3 Değerlendirme metrikleri

Geliştirilen modellerin performansı, Dice ve Jaccard katsayıları kullanılarak değerlendirilmiştir. Dice katsayısı, tahmin edilen bölütleme maskeleri ile gerçek (ground-truth) maskeler arasındaki örtüşme derecesini ölçen bir performans metriğidir. Tıbbi görüntü bölütleme çalışmalarında yaygın olarak kullanılan bu metrik, 0 ile 1 arasında değer almakta olup, Denklem 4.1'de verilen şekilde hesaplanmaktadır.

$$\text{Dice katsayısı} = \frac{2|A \cap B|}{|A| + |B|} \tag{4.1}$$

Burada A tahmin edilen maskeyi, B ise gerçek maskeyi temsil etmektedir. Dice katsayısı, ACDC yarışması kapsamında sunulan değerlendirme kodları ve medpy kütüphanesi kullanılarak hesaplanmıştır.

Geliştirilen modellerin performansı ayrıca, Jaccard katsayısı kullanılarak da değerlendirilmiştir. Jaccard katsayısı, Dice katsayısı benzer şekilde tıbbi görüntü bölütleme çalışmalarında yaygın olarak kullanılan bir metriktir ve 0 ile 1 arasında değer almaktadır. Jaccard skoru, Denklem 4.2'de gösterildiği şekilde hesaplanmaktadır.

$$\text{Jaccard katsayısı} = \frac{|A \cap B|}{|A \cup B|} \tag{4.2}$$

Burada A tahmin edilen maskeyi, B ise gerçek maskeyi temsil etmektedir. Jaccard katsayısı, Dice katsayısı kıyasla daha katı bir değerlendirme ölçütü olup, özellikle küçük

örtüşme hatalarına karşı daha duyarlıdır. Bu çalışmada Jaccard katsayısı medpy kütüphanesi kullanılarak hesaplanmıştır.

4.3 Bulgular

4.3.1 Nicel bulgular

Farklı kodlayıcı başlatma stratejilerinin etiket-verimliliğini değerlendirmek amacıyla, ResNet-50 kodlayıcısına sahip bir U-Net modelinin bölütleme performansı; rastgele başlatma, ImageNet ön-eğitimi ve üç farklı öz-denetimli öğrenme yöntemi (BYOL, BYOL+DSA ve BYOL+PSA) kullanılarak karşılaştırılmıştır. Değerlendirmeler, farklı miktarlarda etiketli veri kullanılarak gerçekleştirilen ince ayar senaryoları altında yapılmıştır. Deneyler, iki farklı kardiyak görüntüleme veri seti olan ACDC (MR) ve CHD (BT) üzerinde gerçekleştirilmiş; değerlendirme metrikleri olarak hem Dice katsayısı hem de Jaccard katsayısı kullanılmıştır. ACDC veri seti üzerindeki nicel sonuçlar, tüm kardiyak yapılar için ortalama Dice katsayısı ve ortalama Jaccard katsayısı değerlerinin raporlandığı Çizelge 4.2 ve Çizelge 4.3'te; Sol Ventrikül (LV), Sağ Ventrikül (RV) ve Miyokardiyum (MYO) için sınıfa özgü Dice ve Jaccard katsayısı sonuçları ise Çizelge 4.4–4.9 arasında verilmiştir. CHD veri setine ait ortalama Dice ve Jaccard sonuçları ise sırasıyla Çizelge 4.10 ve Çizelge 4.11'de sunulmaktadır.

Çizelgelerde M, denetimli ince ayar aşamasında kullanılan hasta sayısını ifade etmektedir. Tüm deneyler ACDC ve CHD veri setleri üzerinde 5-katlı çapraz doğrulama kullanılarak gerçekleştirilmiştir. ACDC veri seti için deneyler ayrıca farklı rastgele tohumlarla (seeds) gerçekleştirilen üç tekrar üzerinden yürütülmüştür. Tüm sonuçlar, ortalama ± standart sapma biçiminde raporlanmış; en iyi sonuçlar **kalın**, ikinci en iyi sonuçlar ise *altı çizili* olarak belirtilmiştir.

ACDC veri setinde, hem Dice katsayısı hem de Jaccard katsayısı metrikleri açısından tüm etiketli veri seviyelerinde BYOL+PSA, en yüksek ortalama Dice skorlarını tutarlı biçimde elde etmiş; BYOL+DSA ise ikinci en iyi yöntem olarak öne çıkmıştır. Bu sonuçlar, öz-denetimli öğrenme ile saçılım tabanlı temsillerin birlikte kullanılmasının kardiyak bölütlemede önemli avantajlar sağladığını göstermektedir. Rastgele başlatma stratejisi özellikle düşük etiketli rejimlerde ($M \leq 10$) hem Dice hem de Jaccard değerleri açısından en zayıf performansı sergilerken, ImageNet ön-eğitimi rastgele başlatmaya kıyasla belirli bir iyileşme sağlasa da alan farkı nedeniyle öz-denetimli yaklaşımların gerisinde kalmaktadır.

Standart BYOL yaklaşımı, ince ayar aşamasından önce etiketlenmemiş kardiyak MR verileri üzerinden alan-özgü temsiller öğrenebilmesi sayesinde, tüm etiketleme seviyelerinde rastgele başlatma ve ImageNet ön-eğitim stratejilerine kıyasla daha üstün performans sergilemektedir.

Etiketli veri miktarı arttıkça tüm yöntemler her iki metrik açısından da istikrarlı bir performans artışı sergilemektedir; bu durum, ek denetimli verinin beklenen faydasını yansıtmaktadır. Bununla birlikte, yöntemler arasındaki performans farkının veri rejimlerine bağlı olarak önemli ölçüde değişkenlik gösterdiği gözlemlenmiştir. Etiketli veri miktarının en yüksek olduğu durumda ($M = 80$), yöntemler arasındaki performans farkı oldukça sınırlı kalmış; en iyi ve en zayıf yöntem arasındaki ortalama Dice katsayısı farkı yalnızca %2.2 olarak ölçülmüştür (BYOL+PSA için 0.9111, rastgele başlatma için ise 0.8891). Etiketli verinin en sınırlı olduğu senaryoda ($M = 5$) yöntemler arasındaki performans farkı belirgin hale gelmiştir. Rastgele başlatma ve ImageNet ön-eğitimi yaklaşımları, anlamlı ölçüde daha düşük performans sergilemiş ve yüksek standart sapma değerleri göstermiştir. Bu durum, sınırlı denetim altında söz konusu yöntemlerin kararlılık ve genelleme yeteneklerinin zayıf kaldığına işaret etmektedir. Buna karşılık, saçılım tabanlı öz-denetimli öğrenme yöntemleri (BYOL+DSA ve BYOL+PSA), yüksek etiket verimliliği sergileyerek etiketli verinin son derece sınırlı olduğu durumlarda dahi kararlı ve güçlü bir bölütleme performansı ortaya koymuştur.

Her iki saçılım tabanlı öz-denetimli öğrenme yöntemi de standart BYOL yaklaşımını geride bıraksa da, en belirgin performans artışları BYOL+PSA yöntemi ile elde edilmiştir. BYOL+PSA, standart BYOL modeline kıyasla ortalama Dice katsayısında $M = 5$ için %4.66 ve $M = 10$ için %2.11 oranında iyileşme sağlamıştır. Bu kazanımlar, saçılım tabanlı ön-uçların öz-denetimli öğrenme yöntemlerine entegre edilmesinin, az etiketli senaryolarda daha güçlü ve daha güvenilir temsiller üretilmesini mümkün kıldığını göstermektedir. Ayrıca, saçılım tabanlı öz-denetimli öğrenme ön-eğitiminin standart öz-denetimli öğrenme temel yaklaşımına kıyasla sağladığı iyileşmenin istatistiksel anlamlılığını değerlendirmek amacıyla, BYOL ve BYOL+PSA kodlayıcılarının Dice katsayıları arasında eşleştirilmiş t-testi (paired t-test) uygulanmıştır. Yapılan analiz, saçılım tabanlı öz-denetimli öğrenme yaklaşımının sağladığı performans kazanımlarının, tüm etiketli veri miktarları genelinde istatistiksel olarak anlamlı olduğunu doğrulamıştır. Tüm veri rejimleri için elde edilen p-değerleri, %5 anlamlılık düzeyinin ($p < 0{,}05$) oldukça altında kaydedilmiştir. Yapılan analiz sonucunda p-değerleri $\mathrm{M} = 5$ için $\mathrm{p} = 1.5073 \times 10^{-5}$, $\mathrm{M} = 10$ için $\mathrm{p} = 7.7736 \times$

10^{-8}, M = 20 için p = 2.9822×10^{-7}, M = 40 için p = 1.5507×10^{-8} ve M = 80 için p = 5.0855×10^{-10} olarak hesaplanmıştır.

Genel bölütleme kalitesi Çizelge 4.2 ve Çizelge 4.3'te özetlenirken, sınıfa özgü Dice ve Jaccard katsayıları, farklı ön-eğitim stratejilerinin her bir kardiyak yapı üzerindeki etkisini ayrıntılı biçimde ortaya koymaktadır (Çizelge 4.4–4.9). En büyük ve en düzenli kardiyak yapı olan Sol Ventrikül (LV) için elde edilen sonuçlar, tüm yöntemler arasında en yüksek Dice ve Jaccard değerlerine bu sınıfta ulaşıldığını göstermektedir (Çizelge 4.4 ve 4.5). Bu durum, LV bölütleme görevinin diğer yapılara kıyasla doğası gereği daha düşük bir zorluk seviyesine sahip olması ve LV ile onu çevreleyen tek yapı olan MYO arasındaki yüksek kontrast ile açıklanabilir. BYOL+PSA ve BYOL+DSA yöntemleri LV için en iyi performansı sunsa da, standart BYOL yaklaşımına kıyasla elde edilen göreli kazanımların en sınırlı kaldığı sınıf LV olmuştur. Bu bulgu, yüksek kaliteli öz-denetimli öğrenme temsillerinin, anatomik olarak daha basit ve iyi tanımlanmış yapıların bölütlenmesinde zaten oldukça etkili olduğunu ortaya koymaktadır.

Daha zorlu anatomik yapılara geçildiğinde, saçılım tabanlı ön-eğitimin sağladığı avantajların çok daha belirgin hale geldiği gözlemlenmiştir. Sağ Ventrikül (RV), yüksek morfolojik değişkenliğe sahip olması ve düşük kontrastlı arka plan bölgeleriyle komşuluğu nedeniyle bölütlenmesi zor bir yapı olarak bilinmektedir (Çizelge 4.6 ve 4.7). Bu durum, RV sınırlarının LV sınırlarına kıyasla belirgin biçimde daha az ayırt edilebilir olmasına yol açmaktadır. Bu bağlamda, BYOL+PSA yöntemi tüm öz-denetim tabanlı olmayan temel yaklaşımlara kıyasla belirgin ve tutarlı bir üstünlük sergilemekte; özellikle düşük etiketli rejimlerde performans farkı belirgin biçimde artmaktadır. Etiketli verinin en sınırlı olduğu durumda ($M = 5$), BYOL+PSA yöntemi 0.7879 Dice katsayısına ulaşmış ve rastgele başlatmaya (0.2872) kıyasla %50.7 oranında bir iyileşme sağlamıştır. Bu bulgular, sınırlı denetim koşulları altında anatomik olarak karmaşık yapıların bölütlenmesinde, etkili bir ön-eğitimin taşıdığı önemi açık biçimde ortaya koymaktadır.

Saçılım tabanlı yöntemlerin göreli performans kazanımlarının en belirgin olduğu durum, miyokardiyum (MYO) bölütleme görevidir (Çizelge 4.8 ve 4.9). Miyokardiyal duvarın ince yapısı nedeniyle bu görev, küçük sınır hatalarına karşı oldukça hassas olup, önerilen yöntemin sağladığı iyileşmeler bu zorlu senaryoda daha net biçimde ortaya çıkmaktadır. MYO bölütleme görevi, hem endokardiyal hem de epikardiyal sınırların doğru biçimde belirlenmesini gerektirdiği için doğası gereği daha zorludur; buna karşılık LV ve RV bölütleme görevleri yalnızca tek bir tek bir kapalı sınırın tanımlanmasını içermektedir.

$M = 5$ durumunda, saçılım ile güçlendirilmiş öz-denetimli öğrenme yöntemleri, standart BYOL yaklaşımına kıyasla en yüksek performans artışını sağlamıştır. Bu bağlamda, BYOL+PSA yöntemi 0.7898 Dice katsayısına ulaşarak, standart BYOL yaklaşımına (0.7338) göre %5.6 oranında bir iyileşme elde etmiştir. Bu sonuç, saçılım tabanlı ön-uçların (DSA ve PSA) öz-denetimli öğrenme yöntemlerine entegre edilmesinin, ince duvarlı ve karmaşık anatomik bölgelerde çok daha doğru sınır tespiti yapabilen nitelikli öznitelik temsilleri oluşturduğu hipotezini kanıtlar niteliktedir. Özetle, BYOL+PSA tüm sınıflar genelinde tutarlı bir biçimde en yüksek performansı sergilese de, üstünlüğü özellikle bölütleme karmaşıklığının ve yapısal değişkenliğin en yüksek olduğu RV ve MYO sınıflarında daha belirgin hale gelmektedir.

Çizelge 4.2: ACDC veri setinde, farklı kodlayıcı başlatma stratejilerine sahip U-Net modellerinin ortalama Dice katsayısı değerlerinin karşılaştırılması.

Ön-Eğitim	M=5	M=10	M=20	M=40	M=80
Random Init.	0.2949 ± 0.1364	0.6054 ± 0.0941	0.7955 ± 0.0291	0.8586 ± 0.0206	0.8891 ± 0.0075
ImageNet	0.5763 ± 0.1491	0.7993 ± 0.0280	0.8579 ± 0.0092	0.8825 ± 0.0056	0.8989 ± 0.0037
BYOL	0.7857 ± 0.0263	0.8435 ± 0.0097	0.8728 ± 0.0039	0.8863 ± 0.0039	0.8986 ± 0.0026
BYOL+DSA	0.8172 ± 0.0172	0.8596 ± 0.0107	0.8830 ± 0.0051	0.8971 ± 0.0053	0.9082 ± 0.0020
BYOL+PSA	**0.8273 ± 0.0138**	**0.8646 ± 0.0096**	**0.8871 ± 0.0071**	**0.9010 ± 0.0045**	**0.9111 ± 0.0022**

Çizelge 4.3: ACDC veri setinde, farklı kodlayıcı başlatma stratejilerine sahip U-Net modellerinin ortalama Jaccard katsayısı değerlerinin karşılaştırılması.

Ön-Eğitim	M=5	M=10	M=20	M=40	M=80
Random Init.	0.2001 ± 0.1014	0.4609 ± 0.0901	0.6731 ± 0.0375	0.7591 ± 0.0283	0.8038 ± 0.0115
ImageNet	0.4519 ± 0.1256	0.6755 ± 0.0369	0.7565 ± 0.0132	0.7934 ± 0.0086	0.8191 ± 0.0058
BYOL	0.6581 ± 0.0336	0.7356 ± 0.0132	0.7786 ± 0.0058	0.7994 ± 0.0059	0.8188 ± 0.0041
BYOL+DSA	0.7012 ± 0.0218	0.7590 ± 0.0152	0.7945 ± 0.0077	0.8166 ± 0.0082	0.8344 ± 0.0031
BYOL+PSA	**0.7140 ± 0.0181**	**0.7670 ± 0.0139**	**0.8011 ± 0.0105**	**0.8230 ± 0.0067**	**0.8391 ± 0.0035**

Çizelge 4.4: ACDC veri setinde, sol ventrikül (LV) bölütlemesi için farklı kodlayıcı başlatma stratejilerine sahip U-Net modellerinin Dice katsayısı değerlerinin karşılaştırılması.

“	M=5	M=10	M=20	M=40	M=80
Random Init.	0.4449 ± 0.1543	0.7008 ± 0.1115	0.8688 ± 0.0232	0.9141 ± 0.0118	0.9316 ± 0.0044
ImageNet	0.7679 ± 0.0655	0.8657 ± 0.0241	0.9094 ± 0.0091	0.9256 ± 0.0059	0.9370 ± 0.0034
BYOL	0.8725 ± 0.0274	0.9066 ± 0.0060	0.9235 ± 0.0057	0.9316 ± 0.0040	0.9380 ± 0.0023
BYOL+DSA	0.9003 ± 0.0116	0.9186 ± 0.0098	0.9311 ± 0.0043	0.9382 ± 0.0044	0.9450 ± 0.0022
BYOL+PSA	**0.9041 ± 0.0096**	**0.9227 ± 0.0067**	**0.9336 ± 0.0059**	**0.9420 ± 0.0031**	**0.9472 ± 0.0011**

Çizelge 4.5: ACDC veri setinde, sol ventrikül (LV) bölütlemesi için farklı kodlayıcı başlatma stratejilerine sahip U-Net modellerinin Jaccard katsayısı değerlerinin karşılaştırılması.

Ön-Eğitim	M=5	M=10	M=20	M=40	M=80
Random Init.	0.3122 ± 0.1271	0.5658 ± 0.1149	0.7734 ± 0.0341	0.8434 ± 0.0187	0.8725 ± 0.0073
ImageNet	0.6341 ± 0.0848	0.7684 ± 0.0357	0.8361 ± 0.0137	0.8625 ± 0.0097	0.8819 ± 0.0057
BYOL	0.7782 ± 0.0390	0.8309 ± 0.0097	0.8587 ± 0.0095	0.8725 ± 0.0070	0.8837 ± 0.0040
BYOL+DSA	0.8213 ± 0.0178	0.8509 ± 0.0158	0.8720 ± 0.0072	0.8843 ± 0.0076	0.8962 ± 0.0038
BYOL+PSA	**0.8274 ± 0.0149**	**0.8576 ± 0.0112**	**0.8764 ± 0.0102**	**0.8909 ± 0.0052**	**0.9001 ± 0.0019**

Çizelge 4.6: ACDC veri setinde, sağ ventrikül (RV) bölütlemesi için farklı kodlayıcı başlatma stratejilerine sahip U-Net modellerinin Dice katsayısı değerlerinin karşılaştırılması.

Ön-Eğitim	M=5	M=10	M=20	M=40	M=80
Random Init.	0.2872 ± 0.1979	0.5960 ± 0.1128	0.7957 ± 0.0296	0.8560 ± 0.0241	0.8921 ± 0.0091
ImageNet	0.5720 ± 0.2460	0.8023 ± 0.0376	0.8592 ± 0.0099	0.8858 ± 0.0055	0.9030 ± 0.0042
BYOL	0.7509 ± 0.0391	0.8244 ± 0.0190	0.8664 ± 0.0062	0.8825 ± 0.0082	0.8995 ± 0.0035
BYOL+DSA	0.7688 ± 0.0330	0.8392 ± 0.0167	0.8710 ± 0.0091	0.8915 ± 0.0102	0.9062 ± 0.0028
BYOL+PSA	**0.7879 ± 0.0272**	**0.8432 ± 0.0175**	**0.8771 ± 0.0118**	**0.8958 ± 0.0088**	**0.9097 ± 0.0046**

Çizelge 4.7: ACDC veri setinde, sağ ventrikül (RV) bölütlemesi için farklı kodlayıcı başlatma stratejilerine sahip U-Net modellerinin Jaccard katsayısı değerlerinin karşılaştırması.

Ön-Eğitim	M=5	M=10	M=20	M=40	M=80
Random Init.	0.1930 ± 0.1379	0.4533 ± 0.1045	0.6747 ± 0.0376	0.7566 ± 0.0327	0.8082 ± 0.0139
ImageNet	0.4424 ± 0.1989	0.6790 ± 0.0474	0.7578 ± 0.0137	0.7978 ± 0.0086	0.8253 ± 0.0067
BYOL	0.6123 ± 0.0464	0.7085 ± 0.0248	0.7688 ± 0.0087	0.7936 ± 0.0122	0.8201 ± 0.0055
BYOL+DSA	0.6374 ± 0.0378	0.7283 ± 0.0235	0.7759 ± 0.0131	0.8080 ± 0.0157	0.8312 ± 0.0042
BYOL+PSA	**0.6597 ± 0.0338**	**0.7359 ± 0.0241**	**0.7858 ± 0.0158**	**0.8147 ± 0.0131**	**0.8369 ± 0.0072**

Çizelge 4.8: ACDC veri setinde, miyokardiyum (MYO) bölütlemesi için farklı kodlayıcı başlatma stratejilerine sahip U-Net modellerinin Dice katsayısı değerlerinin karşılaştırılması.

Ön-Eğitim	M=5	M=10	M=20	M=40	M=80
Random Init.	0.1526 ± 0.1766	0.5192 ± 0.0841	0.7220 ± 0.0418	0.8057 ± 0.0270	0.8437 ± 0.0104
ImageNet	0.3890 ± 0.2831	0.7300 ± 0.0333	0.8052 ± 0.0138	0.8363 ± 0.0088	0.8567 ± 0.0055
BYOL	0.7338 ± 0.0306	0.7996 ± 0.0077	0.8284 ± 0.0058	0.8448 ± 0.0047	0.8584 ± 0.0039
BYOL+DSA	0.7824 ± 0.0155	0.8210 ± 0.0117	0.8469 ± 0.0051	0.8616 ± 0.0044	0.8734 ± 0.0037
BYOL+PSA	**0.7898 ± 0.0120**	**0.8278 ± 0.0122**	**0.8507 ± 0.0081**	**0.8652 ± 0.0051**	**0.8763 ± 0.0023**

Çizelge 4.9: ACDC veri setinde, Miyokardiyum (MYO) bölütlemesi için farklı kodlayıcı başlatma stratejilerine sahip U-Net modellerinin Jaccard katsayısı değerlerinin karşılaştırılması.

Ön-Eğitim	M=5	M=10	M=20	M=40	M=80
Random Init.	0.0952 ± 0.1136	0.3635 ± 0.0742	0.5713 ± 0.0494	0.6773 ± 0.0350	0.7306 ± 0.0152
ImageNet	0.2792 ± 0.2131	0.5791 ± 0.0406	0.6755 ± 0.0190	0.7197 ± 0.0127	0.7501 ± 0.0083
BYOL	0.5838 ± 0.0367	0.6675 ± 0.0104	0.7082 ± 0.0082	0.7322 ± 0.0068	0.7525 ± 0.0059
BYOL+DSA	0.6450 ± 0.0204	0.6980 ± 0.0152	0.7354 ± 0.0075	0.7576 ± 0.0067	0.7758 ± 0.0055
BYOL+PSA	**0.6548 ± 0.0159**	**0.7075 ± 0.0168**	**0.7410 ± 0.0120**	**0.7634 ± 0.0074**	**0.7804 ± 0.0036**

CHD veri seti üzerinde gerçekleştirilen deneylerin nicel sonuçları, farklı kodlayıcı başlatma stratejilerinin etiket-verimliliği üzerindeki etkilerini ortaya koymaktadır (Çizelge 4.10 ve Çizelge 4.11). Dice ve Jaccard katsayıları birlikte değerlendirildiğinde, özellikle etiketli hasta sayısının son derece sınırlı olduğu $M = 2$ ve $M = 5$ senaryolarında öz-denetimli öğrenme tabanlı başlatma stratejilerinin genel olarak rastgele başlatma ve ImageNet ön-eğitimine kıyasla daha yüksek ve daha kararlı performanslar sunduğu görülmektedir. En kısıtlı etiket senaryosu olan $M = 2$ durumunda, BYOL + PSA yaklaşımı Dice katsayısı açısından en yüksek performansı elde ederken, Jaccard metriğinde de öz-denetimli öğrenme tabanlı başlatma stratejilerinin belirgin bir üstünlük sağladığı gözlemlenmiştir. $M = 5$ senaryosunda ise Jaccard katsayısı açısından en yüksek değerlerin

BYOL ve BYOL + DSA yaklaşımları tarafından elde edildiği, buna karşın BYOL + PSA'nın Dice katsayısı bakımından rekabetçi bir performans sergilediği görülmektedir. Etiketli hasta sayısının artmasıyla birlikte, güçlü başlatma stratejileri arasındaki performans farklarının giderek azaldığı ve bazı senaryolarda yöntemlerin göreli sıralamalarının değiştiği dikkat çekmektedir. Yüksek etiketli senaryolarda gözlemlenen farkların görece küçük olması, denetimli bilginin yeterli hale gelmesiyle birlikte farklı başlatma stratejilerinin benzer performans seviyelerine yakınsadığını göstermektedir. Genel olarak elde edilen bulgular, saçılım tabanlı öz-denetimli öğrenme yaklaşımlarının özellikle az etiketli ve yüksek anatomik değişkenliğe sahip CHD veri setinde yüksek etiket-verimliliği sağladığını; buna karşılık etiketli veri miktarının artmasıyla birlikte klasik ön-eğitim stratejilerinin de karşılaştırılabilir performanslara ulaşabildiğini ortaya koymaktadır.

Çizelge 4.10: CHD veri setinde, farklı kodlayıcı başlatma stratejilerine sahip U-Net modellerinin ortalama Dice katsayısı değerlerinin karşılaştırılması.

Ön-Eğitim	M=2	M=5	M=10	M=20	M=40
Random Init.	0.2182 ± 0.0748	0.2971 ± 0.0564	0.5149 ± 0.0392	0.5764 ± 0.0158	0.6395 ± 0.0064
ImageNet	0.3027 ± 0.0696	0.4567 ± 0.0372	0.5901 ± 0.0207	0.6372 ± 0.0116	**0.6982 ± 0.0037**
BYOL	0.3493 ± 0.0523	0.4692 ± 0.0785	0.5842 ± 0.0178	0.6440 ± 0.0118	0.6904 ± 0.0048
BYOL+DSA	0.3286 ± 0.0688	**0.4733 ± 0.0795**	0.5654 ± 0.0571	**0.6498 ± 0.0079**	0.6959 ± 0.0073
BYOL+PSA	**0.3547 ± 0.0637**	0.4673 ± 0.0707	**0.5940 ± 0.0130**	0.6451 ± 0.0159	0.6873 ± 0.0090

Çizelge 4.11: CHD veri setinde, farklı kodlayıcı başlatma stratejilerine sahip U-Net modellerinin ortalama Jaccard katsayısı değerlerinin karşılaştırılması.

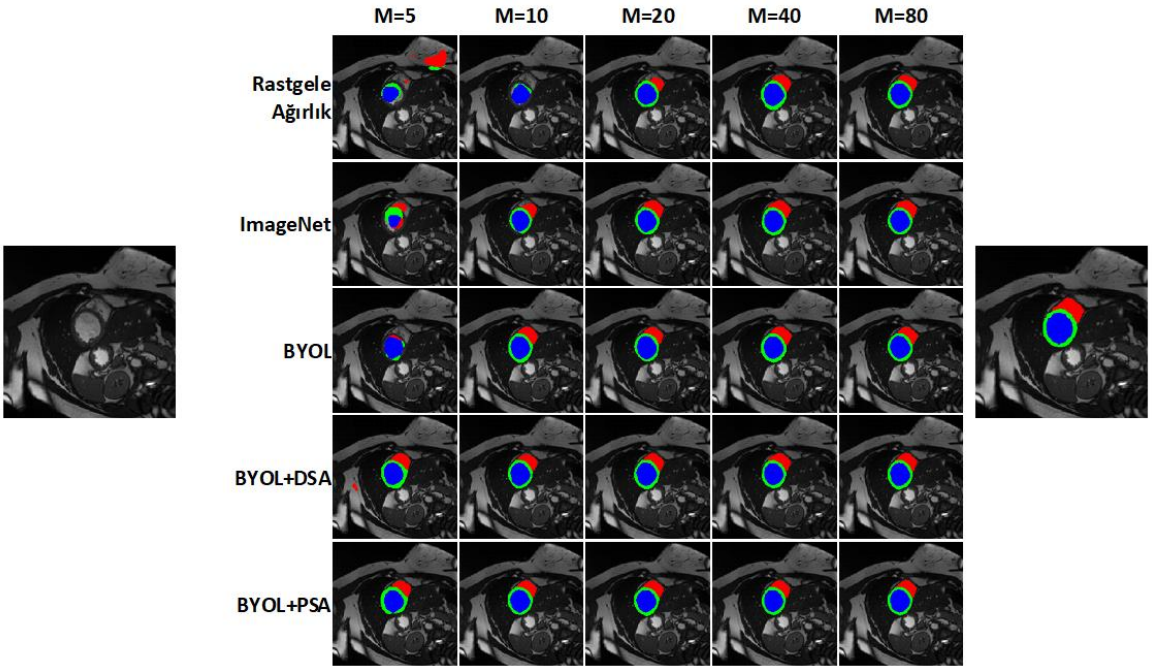
Ön-Eğitim	M=2	M=5	M=10	M=20	M=40
Random Init.	0.1745 ± 0.0743	0.2555 ± 0.0583	0.4455 ± 0.0223	0.5056 ± 0.0187	0.5680 ± 0.0093
ImageNet	0.2596 ± 0.0653	0.4036 ± 0.0321	0.5173 ± 0.0191	0.5624 ± 0.0109	**0.6236 ± 0.0035**
BYOL	**0.2994 ± 0.0484**	**0.4083 ± 0.0665**	0.5100 ± 0.0179	0.5690 ± 0.0119	0.6164 ± 0.0035
BYOL+DSA	0.2721 ± 0.0598	0.4071 ± 0.0694	0.5002 ± 0.0443	**0.5749 ± 0.0082**	0.6217 ± 0.0070
BYOL+PSA	0.2966 ± 0.0559	0.4001 ± 0.0620	**0.5202 ± 0.0126**	0.5713 ± 0.0155	0.6130 ± 0.0086

4.3.2 Nitel bulgular

Nicel değerlendirmelerde elde edilen bulguların görsel olarak daha sezgisel biçimde yorumlanabilmesi amacıyla, farklı ön-eğitim ve başlatma stratejileriyle eğitilmiş modellerin ürettiği bölütleme maskeleri nitel olarak karşılaştırılmıştır. Bu kapsamda, Şekil 4.3'te, farklı etiketli veri seviyelerinde elde edilen örnek bölütleme sonuçları gösterilmektedir. Nitel karşılaştırmaların temsil edici olmasını sağlamak amacıyla, her bir yöntem ve etiketleme seviyesi için test kümesinde elde edilen Dice katsayısı ilgili yöntemin 5-katlı çapraz doğrulama ve 3 tekrar sonucunda hesaplanan ortalama Dice katsayısına en yakın olan model kontrol noktası (checkpoint) seçilmiştir. Bu yaklaşım, sunulan görsellerin şans eseri seçilmiş uç örnekler (cherry-picked) yerine, modellerin genel karakteristiğini yansıtan temsilci örnekler olmasını sağlamaktadır.

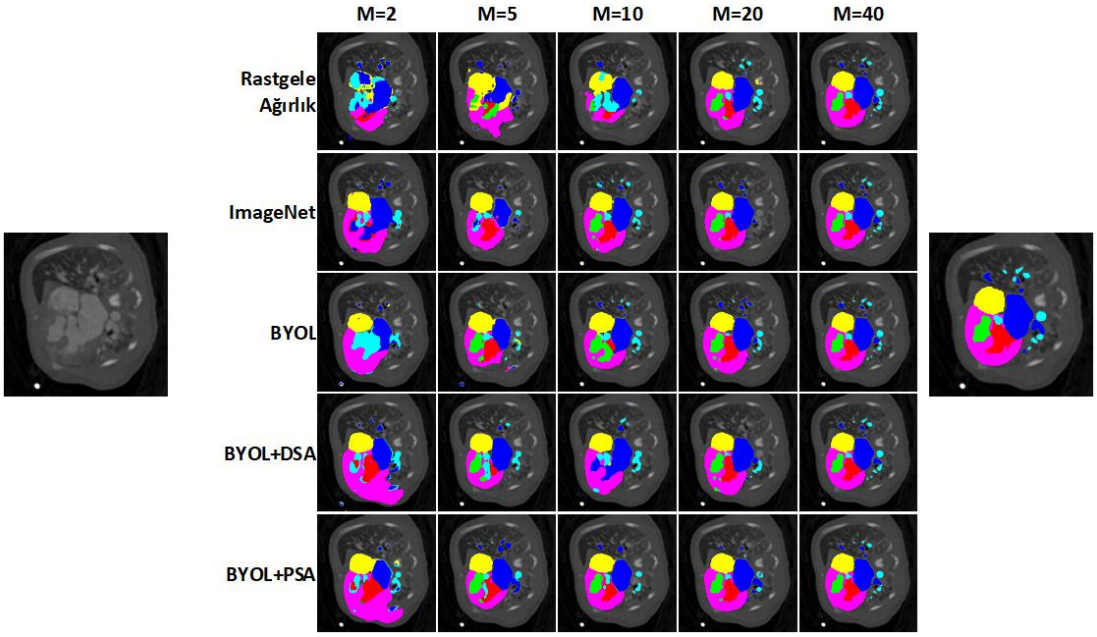
Görsel analiz için, ACDC veri setinde yer alan dilate kardiyomiyopati (DCM) tanısına sahip bir hastadan seçilen, orta-ventriküler düzeyde ve end-diyastolik (ED) faza ait bir MR kesiti kullanılmıştır. Bu seçimin temel gerekçesi, DCM vakalarında sol ventrikül boşluğunun belirgin biçimde genişlemesi ve miyokard duvarının genel olarak incelmesi nedeniyle, özellikle MYO sınırlarının belirsiz hale gelmesidir. Bu durum, söz konusu kesiti sınır lokalizasyonu açısından oldukça zorlayıcı kılmakta ve farklı ön-eğitim stratejilerinin temsil gücünü karşılaştırmak için uygun bir test örneği sunmaktadır.

Şekil 4.3'te görüldüğü üzere, saçılım-tabanlı öz-denetimli öğrenme yaklaşımları, özellikle etiketli verinin en kısıtlı olduğu düşük veri rejiminde ($M = 5$) diğer yöntemlere kıyasla daha tutarlı ve anatomik olarak daha anlamlı bölütleme sonuçları üretmektedir. Rastgele başlatma ve ImageNet ön-eğitimi ile başlatılan modellerde, miyokard sınırlarında kopukluklar, aşırı ya da eksik bölütleme gibi hatalar belirgin biçimde gözlemlenirken; standart BYOL yaklaşımı bu hataları kısmen azaltmakla birlikte, özellikle ince miyokard bölgelerinde sınır kararsızlığı sergilemektedir. Buna karşılık, BYOL+DSA ve özellikle BYOL+PSA yöntemleri, sol ventrikül, sağ ventrikül ve miyokard yapıları arasında daha net sınırlar oluşturmakta ve anatomik bütünlüğü daha başarılı biçimde korumaktadır. PSA tabanlı yapı, miyokardın inceldiği bölgelerde dahi daha düzgün ve süreklilik gösteren sınırlar üretmiş; bu durum, saçılım temsillerinin sunduğu öteleme değişmezliği ve deformasyon kararlılığı gibi özelliklerin, öz-denetimli öğrenme sürecinde öğrenilen temsillere doğrudan yansıdığını göstermektedir.



Şekil 4.3: ACDC veri seti için örnek bölütleme maskesi görselleştirmesi.

CHD veri seti üzerindeki bölütleme performansını nitel olarak değerlendirmek amacıyla, farklı başlatma stratejileri kullanılarak ve farklı etiketli hasta sayıları ($M \in \{2, 5, 10, 20, 40\}$) altında elde edilen temsili bölütleme sonuçları Şekil 4.4'te sunulmaktadır. Şekil 4.4'te görüldüğü üzere, rastgele başlatma yöntemi, özellikle etiketli verinin son derece sınırlı olduğu senaryolarda, parçalı ve anatomik olarak tutarsız bölütleme çıktıları üretmektedir. Buna karşılık, ön-eğitimli modellerin daha bütüncül, kararlı ve anatomik açıdan daha tutarlı tahminler sunduğu gözlemlenmektedir. Özellikle, önerilen saçılım tabanlı öz-denetimli öğrenme yaklaşımlarının, $M = 2$ ve $M = 5$gibi düşük etiketli senaryolarda, daha kararlı ve anatomik olarak daha makul bölütleme sonuçları ürettiği dikkat çekmektedir. Bu gözlemler, önceki bölümde sunulan nicel performans sonuçlarıyla da uyumludur. Etiketli hasta sayısının artmasıyla birlikte, güçlü başlatma stratejileri arasındaki görsel performans farklarının azaldığı ve sonuçların giderek daha benzer hale geldiği görülmektedir.



Şekil 4.4: CHD veri seti için örnek bölütleme maskesi görselleştirmesi.

4.4 Tartışma

Bu çalışmada, sınırlı etiketli veri koşulları altında kardiyak görüntü bölütleme görevleri için saçılım tabanlı öznitelik çıkarıcıların öz-denetimli öğrenme yöntemleriyle entegrasyonunun etkinliğini iki farklı veri seti (ACDC ve CHD) üzerinden incelenmiştir. Elde edilen sonuçlar, saçılım tabanlı öz-denetimli ön-eğitimin; rastgele başlatma, ImageNet tabanlı denetimli ön-eğitim ve standart BYOL yaklaşımlarına kıyasla farklı hasta grupları ve belirgin anatomik varyasyonlar içeren veri setlerinde dahi tutarlı ve belirgin üstünlükler

sağladığını göstermektedir. Bu bulgular, sabit veya parametrik saçılım ön-uçlarının, bölütleme görevlerine yüksek aktarılabilirlik sergileyen, kararlı ve etiket açısından verimli temsiller ürettiğini ortaya koyarak çalışmanın temel katkısını açık biçimde ortaya koymaktadır.

Elde edilen bulgular, etiketli veri kümelerinin doğası gereği sınırlı olduğu tıbbi görüntü analizi alanında öz-denetimli ön-eğitimin önemli bir potansiyele sahip olduğunu göstermektedir. Öz-denetimli öğrenme temelli yaklaşımlarla (BYOL, BYOL+DSA ve BYOL+PSA) başlatılan bölütleme modellerinin, özellikle etiketli verinin sınırlı olduğu senaryolarda ve birçok anatomik sınıf için, rastgele başlatma ve ImageNet ön-eğitimi gibi klasik başlatma stratejilerine kıyasla daha tutarlı ve istikrarlı performanslar sunduğu gözlemlenmiştir. Buna karşın, etiketli veri miktarının arttığı senaryolarda ImageNet ön-eğitimi ile başlatılan modellerin de rekabetçi ve bazı durumlarda daha yüksek performanslara ulaşabildiği görülmektedir. Rastgele veya ImageNet ön-eğitimli ağırlıklarla başlatılan kodlayıcıların düşük veri rejimlerinde daha yüksek varyans sergilemesi, bu yaklaşımların sınırlı denetim altında genelleme açısından kırılgan olabileceğine işaret etmektedir. Her ne kadar ImageNet ön-eğitimi rastgele başlatmaya kıyasla daha güçlü bir başlangıç noktası sunsa da, kaynak ve hedef alanlar arasındaki belirgin alan uyumsuzluğu (domain mismatch) nedeniyle tıbbi görüntüleme uygulamalarındaki etkinliği sınırlı kalmaktadır [85]. Buna karşılık, öz-denetimli öğrenme yaklaşımlarının temsilleri doğrudan etiketlenmemiş kardiyak görüntüler üzerinden öğrenmesi, alan kaymasını azaltarak özniteliklerin hedef alanın yapısal özelliklerine daha iyi uyarlanmasına olanak sağlamaktadır.

Değerlendirme sürecinde Dice katsayısına ek olarak kullanılan Jaccard metriği, yöntemler arasındaki performans farklarını daha katı bir örtüşme ölçütü üzerinden analiz etme imkanı sunmuştur. Jaccard metriğinin, küçük yapılar ve sınır bölgelerinde hatalara daha duyarlı olması nedeniyle, saçılım tabanlı yöntemlerin sağladığı yapısal kararlılığın özellikle anlamlı olduğu gözlemlenmiştir.

Karşılaştırmalı öz-denetimli öğrenme yöntemleri doğal görüntü alanında güçlü performanslar sergilemiş olsa da, tıbbi görüntülemenin doğasından kaynaklanan bazı kritik zorluklar nedeniyle bu çalışmada karşılaştırmalı olmayan BYOL yöntemi tercih edilmiştir. Tıbbi görüntü alandaki temel sınırlamalardan biri, güvenilir negatif örnek çiftlerinin tanımlanmasının güçlüğüdür. Doğal görüntüler yüksek görsel çeşitlilik sergilerken, tıbbi görüntüler genellikle homojen bir yapıya sahiptir ve patolojik farklılıklar çoğu zaman

oldukça ince düzeydedir. Bu durum, negatif örneklerin güvenilir biçimde belirlenmesini zorlaştırmakta ve yanlış negatif eşleşmelerin oluşma riskini önemli ölçüde artırmaktadır. Bu nedenle, negatif örnek kullanımına ihtiyaç duymayan yapısı sayesinde BYOL, çalışmada tercih edilen öz-denetimli öğrenme yöntemi olarak seçilmiştir.

Çalışmanın temel bulgularından biri, saçılım ağlarının öz-denetimli öğrenme sürecine entegre edilmesiyle standart BYOL yaklaşımına kıyasla istatistiksel olarak anlamlı bir performans artışı elde edilmesidir. Bu sonuç, etiketli verinin sınırlı olduğu tıbbi görüntüleme görevlerinde öznitelik çıkarım aşamasına güçlü bir matematiksel önsel bilginin dahil edilmesinin önemini açık biçimde ortaya koymaktadır. BYOL yaklaşımı, farklı artırımlarla elde edilen görünümler arasındaki uyumu maksimize ederek temsil öğrenimi gerçekleştirse de, öğrenilen temsiller üzerinde yapısal bir kısıtlama dayatmamaktadır. Buna karşılık DSA, öteleme değişmezliği ve küçük deformasyonlara karşı kararlılığı garanti eden matematiksel bir çerçeveye dayanmaktadır. Bu özellikler, küçük konumsal kaymaların ve hasta bazlı anatomik farklılıkların sıkça görüldüğü kardiyak görüntülerinde, altta yatan anatomik yapıların daha kararlı ve doğru biçimde tanımlanmasına olanak sağlamaktadır.

Her iki saçılım yaklaşımı da dalgacık temsillerinin sağladığı kararlılık özelliklerini paylaşsa da, BYOL+PSA yönteminin BYOL+DSA'ya kıyasla tutarlı biçimde daha üstün performans sergilemesi, PSA'da kullanılan öğrenilebilir filtrelerden kaynaklanmaktadır. DSA, önceden tanımlanmış sabit bir filtre bankasına dayanarak kararlı ancak görece genel temsiller üretirken; PSA, dalgacık parametrelerini eğitim süreci boyunca optimize ederek saçılım temsillerinin göreve özgü hale gelmesini sağlamaktadır. Bu sayede, temsiller altta yatan anatomik yapılarla daha iyi uyum göstermekte ve daha ayırt edici bir temsil gücü elde edilmektedir.

Saçılım tabanlı bu temsilin etkinliği, sınıfa özgü bölütleme sonuçlarında nicel olarak açık biçimde gözlemlenmektedir. En belirgin kazanımlar, standart BYOL yaklaşımına kıyasla en yüksek göreli performans artışının elde edildiği, oldukça zorlayıcı MYO bölütleme görevinde ortaya çıkmıştır. Bu bulgu, saçılım ön-uç tarafından kazandırılan yapısal farkındalık ve sınır duyarlılığı özelliklerinin, ince ve karmaşık miyokardiyal sınırların doğru biçimde ayrıştırılmasında özellikle etkili olduğunu güçlü biçimde desteklemektedir. Ayrıca BYOL+PSA yönteminin anatomik sınıfların genelinde tutarlı üstün performans sergilemesi, değerlendirilen yöntemler arasında en etkili başlatma stratejisi olduğunu doğrulamaktadır.

CHD veri seti üzerinde elde edilen sonuçlar, saçılım tabanlı öz-denetimli başlatma stratejilerinin özellikle az etiketli senaryolarda anlamlı avantajlar sunduğunu ortaya koymaktadır. Bu veri setinde, sınırlı etiketleme koşulları altında BYOL+DSA ve BYOL+PSA yöntemleri Dice metriği açısından, rastgele başlatma ve ImageNet ön-eğitimli modellere kıyasla daha başarılı sonuçlar üretmiştir. Buna karşılık, etiket miktarının arttığı senaryolarda ImageNet ön-eğitiminin hem Dice hem de Jaccard metrikleri açısından üstün performans sergilemesi, denetimli ön-eğitimin yeterli etiketli veri mevcut olduğunda daha ayrıntılı ve sınır hassasiyeti yüksek temsiller öğrenebildiğini göstermektedir. Bu bulgular, saçılım tabanlı öz-denetimli başlatmanın temel avantajının etiket-verimliliği olduğunu ve özellikle veri kısıtlı klinik senaryolar için uygun bir başlangıç stratejisi sunduğunu desteklemektedir. Dice metriğinde gözlemlenen kazanımların Jaccard metriğinde aynı ölçüde yansımaması, Jaccard'ın küçük sınır hatalarına daha duyarlı ve daha katı bir örtüşme ölçütü olmasından kaynaklanmaktadır. Saçılım tabanlı temsiller, sınırlı veri koşullarında anatomik yapıların genel formunu ve bütünlüğünü başarıyla yakalarken, yüksek etiketli senaryolarda ImageNet ön-eğitimiyle başlatılan modellerin daha ince sınır detaylarını öğrenme konusunda avantaj sağladığı gözlemlenmiştir.

Bu çalışmada elde edilen umut verici sonuçlara rağmen, bazı sınırlılıklar mevcuttur. İlk olarak, gerçekleştirilen deneyler yalnızca tek bir anatomik bölge (kalp) ile sınırlıdır. Saçılım tabanlı bileşenlerin farklı anatomik yapılar üzerindeki dayanıklılığının değerlendirilmesi, yöntemin daha geniş ölçekte genellenebilirliğinin ortaya konması açısından önem taşımaktadır. İkinci olarak, saçılım tabanlı dönüşümler, PSA modelinde dahi ölçek ve yönelim sayısı gibi belirli tasarım tercihlerini içermektedir. Bu sabit yapılandırmalar, yöntemin tüm miyokardiyal alt yapılar veya farklı anatomik varyasyonlar üzerinde tam esneklik göstermesini sınırlayabilir. Bu bağlamda, saçılım tabanlı öz-denetimli başlatma stratejilerinin sınır hassasiyetini artırmaya yönelik ek mekanizmalarla (örneğin sınır odaklı kayıp fonksiyonları veya çok-ölçekli denetim) desteklenmesi, Jaccard metriği açısından performansın iyileştirilmesi için gelecek çalışmalar açısından önemli bir araştırma yönü olarak değerlendirilmektedir. Son olarak, bu çalışmada değerlendirme yalnızca iki boyutlu bölütleme ile sınırlandırılmıştır. Önerilen yaklaşımın üç boyutlu mimarilere genişletilmesi, hem ek zorlukları hem de potansiyel performans kazanımlarını ortaya çıkarabilecek önemli bir araştırma yönü olarak değerlendirilmektedir.

5. SONUÇ ve GELECEK ÇALIŞMALAR

Bu tez çalışmasında, tıbbi görüntü bölütleme problemlerinde etiket verisine olan bağımlılığı azaltmayı hedefleyen saçılım-tabanlı bir öz-denetimli öğrenme yaklaşımı önerilmiştir. Önerilen yöntemde, Dalgacık Saçılım Ağları (DSA) ve Parametrik Dalgacık Saçılım Ağları (PSA), karşılaştırmalı olmayan bir öz-denetimli öğrenme yöntemi olan BYOL mimarisine entegre edilmiş; böylece kodlayıcı ağın başlangıç katmanlarında matematiksel olarak temellendirilmiş, yapısal olarak anlamlı ve alan-özgü özniteliklerin öğrenilmesi amaçlanmıştır. Bu sayede, büyük ölçekli ve etiketlenmemiş kardiyak görüntü verisi kullanılarak, manuel etiketleme gereksinimi olmadan anatomik olarak ayırt edici temsil öğrenimi gerçekleştirilmiştir.

Ön-eğitim sürecinde elde edilen bu temsiller, daha sonra bir U-Net mimarisine aktarılmış ve sınırlı miktarda etiketli veri kullanılarak ince ayar aşamasında değerlendirilmiştir. Deneysel sonuçlar, önerilen saçılım-tabanlı öz-denetimli öğrenme yaklaşımının; rastgele başlatma ve ImageNet ön-eğitimi gibi geleneksel başlatma stratejilerine kıyasla anlamlı ve tutarlı performans artışları sağladığını göstermektedir. Ayrıca, standart BYOL yaklaşımı ile karşılaştırıldığında, özellikle sınırlı etiketli veri senaryolarında belirgin üstünlükler elde edilmiştir. Bu bulgu, saçılım dönüşümlerinin sunduğu öteleme değişmezliği ve deformasyon kararlılığı gibi kuramsal özelliklerin, öz-denetimli öğrenme temsillerinin genelleme kabiliyetini artırmada kritik bir rol oynadığını ortaya koymaktadır.

Çalışmanın önemli bir diğer sonucu, PSA tabanlı yaklaşımın, sabit dalgacık filtreleri kullanan DSA'ya kıyasla daha yüksek performans elde etmesidir. PSA mimarisinde dalgacık parametrelerinin eğitim süreci boyunca öğrenilebilir olması, alan-özgü karakteristiklere ve bölütleme görevine daha etkin biçimde uyum sağlamasına olanak tanımıştır. Bu durum, matematiksel olarak yapılandırılmış öncüllerin tamamen sabit tutulması yerine, öğrenilebilir hale getirilmesinin, hem teorik hem de pratik açıdan avantajlı olabileceğini göstermektedir. Elde edilen sonuçlar, özellikle PSA'nın sunduğu bu esnekliğin, öz-denetimli temsil öğreniminde daha ayırt edici ve görev-uyumlu temsiller elde edilmesini sağladığını ortaya koymaktadır.

Bu tez kapsamında sunulan bulgular, saçılım tabanlı öz-denetimli yöntemlerinin potansiyelini açıkça ortaya koymakla birlikte, gelecekte ele alınması gereken çeşitli araştırma yönleri de bulunmaktadır. Öncelikle, bu çalışmada kullanılan veri artırma

stratejileri büyük ölçüde BYOL literatüründe yaygın olarak kullanılan standart dönüşümlerle sınırlıdır. Gelecek çalışmalarda, saçılım yöntemlerinin doku ve ince ayrıntıları yakalamadaki hassasiyeti göz önüne alınarak, temsille daha uyumlu ve alan-özgü veri artırma stratejileri geliştirilebilir. Ayrıca, ön-eğitim aşamasında kullanılan artırmalar ile bölütleme görevine özgü ince ayar aşamasındaki dönüşümler arasındaki uyumun artırılması, temsil-transfer performansını daha da iyileştirebilir.

İkinci olarak, bu tezde U-Net mimarisi, literatürdeki yaygın kullanımı ve karşılaştırılabilirlik açısından temel bir referans model olarak tercih edilmiştir. Ancak, önerilen saçılım-tabanlı öz-denetimli öğrenme yönteminin, nnU-Net [86], DeepLabV3+ [87] veya U-Net++ [88] gibi daha gelişmiş ve adaptif mimarilerle birlikte değerlendirilmesi, yöntemin mimariden bağımsız genellenebilirliğini daha kapsamlı biçimde ortaya koyacaktır. Özellikle otomatik yapılandırma ve hiperparametre optimizasyonu yeteneklerine sahip mimarilerle yapılacak çalışmalar, yaklaşımın pratik klinik senaryolardaki uygulanabilirliğini güçlendirebilir.

Son olarak, bu çalışmanın temel sınırlamalarından biri, deneylerin tek bir anatomik bölge (kalp) ile sınırlandırılmış olmasıdır. Gelecek çalışmalarda, önerilen yaklaşımın farklı anatomik yapılarda değerlendirilmesi, yöntemin genellenilebilirliğini test etmek açısından kritik olacaktır. Ayrıca, mevcut çalışma iki boyutlu dilim tabanlı bir çerçeve üzerine kuruludur. Saçılım tabanlı öz-denetimli öğrenme yöntemlerinin üç boyutlu hacimsel ve dört boyutlu uzamsal-zamansal bölütleme problemlerine genişletilmesi, özellikle kardiyak döngünün zamansal dinamiklerini modellemek açısından önemli yeni araştırma fırsatları sunmaktadır.

Sonuç olarak, bu tez, matematiksel olarak temellendirilmiş saçılım dönüşümlerinin öz-denetimli öğrenme yöntemlerine entegre edilmesinin, etiket-verimli tıbbi görüntü bölütleme alanında güçlü bir araştırma yönü sunduğunu göstermektedir. Gelecekte yapılacak çalışmalarla bu yaklaşımın kapsamı genişletilerek, klinik uygulamalara daha yakın ve genellenebilir çözümler geliştirilmesi mümkün olacaktır.

KAYNAKLAR

[1] **Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., Van Ginneken, B., Madabhushi, A., ... & Summers, R. M.** (2021). A review of deep learning in medical imaging: Imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, 109(5), 820-838.

[2] **Jin, C., Guo, Z., Lin, Y., Luo, L., & Chen, H.** (2023). Label-efficient deep learning in medical image analysis: Challenges and future directions. *arXiv preprint arXiv:2303.12484.*

[3] **Kumari, S., & Singh, P.** (2023). Data efficient deep learning for medical image analysis: A survey. *arXiv preprint arXiv:2310.06557.*

[4] **Upadhyay, A. K., & Bhandari, A. K.** (2024). Advances in deep learning models for resolving medical image segmentation data scarcity problem: a topical review. *Archives of Computational Methods in Engineering*, 31(3), 1701-1719.

[5] **Sultana, F., Sufian, A., & Dutta, P.** (2020). Evolution of image segmentation using deep convolutional neural network: A survey. *Knowledge-Based Systems*, 201, 106062.

[6] **LeCun, Y., Bengio, Y., & Hinton, G.** (2015). Deep learning. *nature*, 521(7553), 436-444.

[7] **Goodfellow, I., Bengio, Y., Courville, A., & Bengio, Y.** (2016). *Deep learning* (Vol. 1, No. 2). Cambridge: MIT press.

[8] **Dumoulin, V., & Visin, F.** (2016). A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285.*

[9] **Dubey, S. R., Singh, S. K., & Chaudhuri, B. B.** (2022). Activation functions in deep learning: A comprehensive survey and benchmark. *Neurocomputing*, 503, 92-108.

[10] **Kauderer-Abrams, E.** (2017). Quantifying translation-invariance in convolutional neural networks. *arXiv preprint arXiv:1801.01450.*

[11] **Noh, H., Hong, S., & Han, B.** (2015). Learning deconvolution network for semantic segmentation. *In Proceedings of the IEEE international conference on computer vision* (pp. 1520-1528).

[12] **Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K.** (2018). Convolutional neural networks: an overview and application in radiology. *Insights into imaging*, 9(4), 611-629.

[13] **Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R.** (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research,* 15(1), 1929-1958.

[14] **Asadi, B., & Jiang, H.** (2020). On approximation capabilities of ReLU activation and softmax output layer in neural networks. *arXiv preprint arXiv:2002.04060.*

[15] **Krizhevsky, A., Sutskever, I., & Hinton, G. E.** (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

[16] **Simonyan, K., & Zisserman, A.** (2014). Very deep convolutional networks for large-

scale image recognition. *arXiv preprint arXiv:1409.1556.*

[17] **Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A.** (2015). Going deeper with convolutions. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1-9).

[18] **He, K., Zhang, X., Ren, S., & Sun, J.** (2016). Deep residual learning for image recognition. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[19] **Weiss, K., Khoshgoftaar, T. M., & Wang, D.** (2016). A survey of transfer learning. *Journal of Big data*, 3(1), 9.

[20] **Pan, S. J., & Yang, Q.** (2009). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345-1359.

[21] **Yosinski, J., Clune, J., Bengio, Y., & Lipson, H.** (2014). How transferable are features in deep neural networks?. *Advances in neural information processing systems*, 27.

[22] **Long, J., Shelhamer, E., & Darrell, T.** (2015). Fully convolutional networks for semantic segmentation. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431-3440).

[23] **Badrinarayanan, V., Kendall, A., & Cipolla, R.** (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence,* 39(12), 2481-2495.

[24] **Ronneberger, O., Fischer, P., & Brox, T.** (2015). U-net: Convolutional networks for biomedical image segmentation. *In International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241).

[25] **Shorten, C., & Khoshgoftaar, T. M.** (2019). A survey on image data augmentation for deep learning. *Journal of big data,* 6(1), 1-48.

[26] **Liu, X., Zhang, F., Hou, Z., Mian, L., Wang, Z., Zhang, J., & Tang, J.** (2021). Self-supervised learning: Generative or contrastive. *IEEE transactions on knowledge and data engineering,* 35(1), 857-876.

[27] **Ericsson, L., Gouk, H., Loy, C. C., & Hospedales, T. M.** (2022). Self-supervised representation learning: Introduction, advances, and challenges. *IEEE Signal Processing Magazine*, 39(3), 42-62.

[28] **Ohri, K., & Kumar, M.** (2021). Review on self-supervised image recognition using deep neural networks. *Knowledge-Based Systems*, 224, 107090.

[29] **Jing, L., & Tian, Y.** (2020). Self-supervised visual feature learning with deep neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 43(11), 4037-4058.

[30] **Bastanlar, Y., & Orhan, S.** (2022). Self-supervised contrastive representation learning in computer vision. *In Artificial Intelligence Annual Volume 2022.* IntechOpen.

[31] **Zhang, R., Isola, P., & Efros, A. A.** (2017). Split-brain autoencoders: Unsupervised learning by cross-channel prediction. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1058-1067).

[32] **Vincent, P., Larochelle, H., Bengio, Y., & Manzagol, P. A**. (2008). Extracting and

composing robust features with denoising autoencoders. *In Proceedings of the 25th international conference on Machine learning* (pp. 1096-1103).

[33] **Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A.** (2016). Context encoders: Feature learning by inpainting. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2536-2544).

[34] **Zhang, R., Isola, P., & Efros, A. A.** (2016). Colorful image colorization. *In European conference on computer vision* (pp. 649-666).

[35] **Doersch, C., Gupta, A., & Efros, A. A.** (2015). Unsupervised visual representation learning by context prediction. *In Proceedings of the IEEE international conference on computer vision* (pp. 1422-1430).

[36] **Noroozi, M., & Favaro, P.** (2016). Unsupervised learning of visual representations by solving jigsaw puzzles. *In European conference on computer vision* (pp. 69-84).

[37] **Gidaris, S., Singh, P., & Komodakis, N.** (2018). Unsupervised representation learning by predicting image rotations. *International Conference on Learning Representations (ICLR)*.

[38] **Caron, M., Bojanowski, P., Joulin, A., & Douze, M.** (2018). Deep clustering for unsupervised learning of visual features. *In Proceedings of the European conference on computer vision (ECCV)* (pp. 132-149).

[39] **Asano, Y. M., Rupprecht, C., & Vedaldi, A.** (2019). Self-labelling via simultaneous clustering and representation learning. *arXiv preprint arXiv:1911.05371*.

[40] **Cuturi, M.** (2013). Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems,* 26.

[41] **Chen, T., Kornblith, S., Norouzi, M., & Hinton, G.** (2020). A simple framework for contrastive learning of visual representations. *In International conference on machine learning* (pp. 1597-1607).

[42] **He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R.** (2020). Momentum contrast for unsupervised visual representation learning. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9729-9738).

[43] **Caron, M., Misra, I., Mairal, J., Goyal, P., Bojanowski, P., & Joulin, A.** (2020). Unsupervised learning of visual features by contrasting cluster assignments. *Advances in neural information processing systems*, 33, 9912-9924.

[44] **Epstein, R.** (2016). The empty brain. *Aeon*, May, 18(3).

[45] **Hadsell, R., Chopra, S., & LeCun, Y.** (2006). Dimensionality reduction by learning an invariant mapping. *In 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06)* (Vol. 2, pp. 1735-1742).

[46] **Schroff, F., Kalenichenko, D., & Philbin, J.** (2015). Facenet: A unified embedding for face recognition and clustering. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 815-823).

[47] **Sohn, K.** (2016). Improved deep metric learning with multi-class n-pair loss objective. *Advances in neural information processing systems,* 29.

[48] **Wu, Z., Xiong, Y., Yu, S. X., & Lin, D.** (2018). Unsupervised feature learning via non-parametric instance discrimination. *In Proceedings of the IEEE conference*

on computer vision and pattern recognition (pp. 3733-3742).

[49] **Gutmann, M., & Hyvärinen, A.** (2010). Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. *In Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 297-304).

[50] **Oord, A. V. D., Li, Y., & Vinyals, O.** (2018). Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748.*

[51] **Alexey, D., Fischer, P., Tobias, J., Springenberg, M. R., & Brox, T.** (2016). Discriminative unsupervised feature learning with exemplar convolutional neural networks. *IEEE TPAMI,* 38(9), 1734-1747.

[52] **Misra, I., & Maaten, L. V. D.** (2020). Self-supervised learning of pretext-invariant representations. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 6707-6717).

[53] **Chen, X., Fan, H., Girshick, R., & He, K.** (2020). Improved baselines with momentum contrastive learning. *arXiv preprint arXiv:2003.04297.*

[54] **Chen, T., Kornblith, S., Swersky, K., Norouzi, M., & Hinton, G. E.** (2020). Big self-supervised models are strong semi-supervised learners. *Advances in neural information processing systems,* 33, 22243-22255.

[55] **Grill, J. B., Strub, F., Altché, F., Tallec, C., Richemond, P., Buchatskaya, E., ... & Valko, M.** (2020). Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33, 21271-21284.

[56] **Chen, X., & He, K.** (2021). Exploring simple siamese representation learning. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 15750-15758).

[57] **Zbontar, J., Jing, L., Misra, I., LeCun, Y., & Deny, S.** (2021). Barlow twins: Self-supervised learning via redundancy reduction. *In International conference on machine learning* (pp. 12310-12320).

[58] **Caron, M., Touvron, H., Misra, I., Jégou, H., Mairal, J., Bojanowski, P., & Joulin, A.** (2021). Emerging properties in self-supervised vision transformers. *In Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9650-9660).

[59] **Dosovitskiy, A.** (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929.*

[60] **Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., ... & Bojanowski, P.** (2023). Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193.*

[61] **Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L.** (2009). Imagenet: A large-scale hierarchical image database. *In 2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255).

[62] **Mallat, S.** (2012). Group invariant scattering. *Communications on Pure and Applied Mathematics,* 65(10), 1331-1398.

[63] **Bruna, J., & Mallat, S.** (2013). Invariant scattering convolution networks. *IEEE transactions on pattern analysis and machine intelligence*, 35(8), 1872-1886.

[64] **Sifre, L., & Mallat, S.** (2013). Rotation, scaling and deformation invariant scattering for texture discrimination. *In Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1233-1240).

[65] **Gauthier, S., Thérien, B., Alsene-Racicot, L., Chaudhary, M., Rish, I., Belilovsky, E., ... & Wolf, G.** (2022). Parametric scattering networks. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 5749-5758).

[66] **Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L.** (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834–848.

[67] **Rani, V., Kumar, M., Gupta, A., Sachdeva, M., Mittal, A., & Kumar, K.** (2024). Self-supervised learning for medical image analysis: a comprehensive review. *Evolving Systems,* 15(4), 1607-1633.

[68] **Akilan, T., Jahan, N., & Zhang, W.** (2025). Self-Supervised Learning for Image Segmentation: A Comprehensive Survey. *arXiv preprint arXiv:2505.13584.*

[69] **Zeng, D., Wu, Y., Hu, X., Xu, X., Yuan, H., Huang, M., ... & Shi, Y.** (2021). Positional contrastive learning for volumetric medical image segmentation. In International conference on medical image computing and computer-assisted intervention (pp. 221-230).

[70] **Punn, N. S., & Agarwal, S.** (2022). BT-Unet: A self-supervised learning framework for biomedical image segmentation using barlow twins with U-net models. *Machine Learning,* 111(12), 4585-4600.

[71] **Kalapos, A., & Gyires-Tóth, B.** (2022). Self-supervised pretraining for 2d medical image segmentation. In European Conference on Computer Vision (pp. 472-484).

[72] **Wang, X., Zhang, R., Shen, C., Kong, T., & Li, L.** (2021). Dense contrastive learning for self-supervised visual pre-training. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3024-3033).

[73] **Xie, Z., Lin, Y., Zhang, Z., Cao, Y., Lin, S., & Hu, H.** (2021). Propagate yourself: Exploring pixel-level consistency for unsupervised visual representation learning. *In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 16684-16693).

[74] **Chaitanya, K., Erdil, E., Karani, N., & Konukoglu, E.** (2020). Contrastive learning of global and local features for medical image segmentation with limited annotations. *Advances in neural information processing systems,* 33, 12546-12558.

[75] **Leterme, H., Polisano, K., Perrier, V., & Alahari, K.** (2022). On the shift invariance of max pooling feature maps in convolutional neural networks. *arXiv preprint arXiv:2209.11740.*

[76] **Oyallon, E., Belilovsky, E., & Zagoruyko, S.** (2017). Scaling the scattering transform: Deep hybrid networks. *In Proceedings of the IEEE international conference on computer vision* (pp. 5618-5627).

[77] **Oyallon, E., Zagoruyko, S., Huang, G., Komodakis, N., Lacoste-Julien, S.,**

Blaschko, M., & Belilovsky, E. (2018). Scattering networks for hybrid representation learning. *IEEE transactions on pattern analysis and machine intelligence,* 41(9), 2208-2221.

[78] **Cotter, F., & Kingsbury, N.** (2019). A learnable scatternet: Locally invariant convolutional layers. *In 2019 IEEE international conference on image processing (ICIP)* (pp. 350-354).

[79] **Kinakh, V., Taran, O., & Voloshynovskiy, S.** (2021). Scatsimclr: self-supervised contrastive learning with pretext task regularization for small-scale datasets. *In Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 1098-1106).

[80] **Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P. A., ... & Jodoin, P. M.** (2018). Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis: is the problem solved?. *IEEE transactions on medical imaging*, 37(11), 2514-2525.

[81] **Xu, X., Wang, T., Shi, Y., Yuan, H., Jia, Q., Huang, M., & Zhuang, J.** (2019). Whole heart and great vessel segmentation in congenital heart disease using deep neural networks and graph matching. *In International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 477-485).

[82] **Li, H., Ouyang, Y., & Wan, X.** (2024). Self-supervised alignment learning for medical image segmentation. *In 2024 IEEE International Symposium on Biomedical Imaging (ISBI)* (pp. 1-5).

[83] **Seince, M., Folgoc, L. L., de Souza, L. A. F., & Angelini, E.** (2024). Dense self-supervised learning for medical image segmentation. *arXiv preprint arXiv:2407.20395.*

[84] **Miron, R., Moisii, C., Dinu, S., & Breaban, M. E.** (2021). Evaluating volumetric and slice-based approaches for COVID-19 detection in chest CTs. *In Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 529-536).

[85] **Wen, Y., Chen, L., Deng, Y., & Zhou, C.** (2021). Rethinking pre-training on medical imaging. *Journal of Visual Communication and Image Representation*, 78, 103145.

[86] **Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J., & Maier-Hein, K. H.** (2021). nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods,* 18(2), 203-211.

[87] **Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H.** (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. *In Proceedings of the European conference on computer vision (ECCV)* (pp. 801-818).

[88] **Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., & Liang, J.** (2018). Unet++: A nested u-net architecture for medical image segmentation. *In International workshop on deep learning in medical image analysis* (pp. 3-11).

ÖZGEÇMİŞ

Ad-Soyad **:** Serdar ALASU

ÖĞRENİM DURUMU:

- **Lisans** **:** 2004, İstanbul Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü
- **Yüksek Lisans** **:** 2018, İnönü Üniversitesi, Bilgisayar Mühendisliği Anabilim Dalı, Bilgisayar Bilimleri Programı
- **Doktora** **:** 2026, İnönü Üniversitesi, Bilgisayar Mühendisliği Anabilim Dalı, Bilgisayar Bilimleri Programı

MESLEKİ DENEYİM:

- 2004 – 2012 : Gaziantep Üniversitesi, Mühendis
- 2012 – Halen : Sağlık Bakanlığı, Mühendis

DOKTORA TEZİNDEN TÜRETİLEN ÇALIŞMALAR

- **Alasu, S., & Talu, M. F.** (2022). Mısır Tohumu Embriyolarının Bölütlenmesinde Tam Evrişimsel Ağ Tabanlı Mimarilerin Tam Bağlı Şartlı Rastgele Alanlar ile Entegrasyonu. Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi, 22(1), 100-111. https://doi.org/10.35414/akufemubid.1013047
- **Alasu, S., Talu, M. F.** (2024). Bilgisayarlı Görüde Öz-Denetimli Öğrenme Yöntemleri Üzerine Bir İnceleme. Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 12(2), 1136-1165. https://doi.org/10.29130/dubited.1201292
- **Alasu, S., Talu, M. F.** (2026). Scattering-Based Self-Supervised Learning for Label-Efficient Cardiac Image Segmentation. Electronics, 15(3), 506. https://doi.org/10.3390/electronics15030506