T.C.
KONYA TEKNİK ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

GNSS SİNYAL KAYIPLARINDA İHA'ların KONUMLANDIRMASI İÇİN DERİN ÖĞRENME TABANLI GÖRSEL ve ATALETSEL VERİ FÜZYONU

Mahmut KARAASLAN

YÜKSEK LİSANS

Bilgisayar Mühendisliği Anabilim Dalı

Haziran-2025
KONYA

TEZ KABUL VE ONAYI

Mahmut Karaaslan tarafından hazırlanan “GNSS Sinyal Kayıplarında İHA'ların Konumlandırılması İçin Derin Öğrenme Tabanlı Görsel ve Ataletsel Veri Füzyonu” adlı tez çalışması 26/06/2025 tarihinde aşağıdaki jüri tarafından oy birliği ile Konya Teknik Üniversitesi Lisansüstü Eğitim Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı’nda YÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Jüri Üyeleri	**İmza**
Başkan Doç. Dr. Mehmet Akif ŞAHMAN Selçuk Üniversitesi	……..
Danışman Doç. Dr. Ersin KAYA Konya Teknik Üniversitesi	……..
Üye Dr. Öğr. Üyesi Sedat KORKMAZ Konya Teknik Üniversitesi	……..

Yukarıdaki sonucu onaylarım.

Prof. Dr. Mevlüt UYAN
Enstitü Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Mahmut KARAASLAN

Tarih: 26/06/2025

ÖZET

YÜKSEK LİSANS

GNSS SİNYAL KAYIPLARINDA İHA'ların KONUMLANDIRMASI İÇİN DERİN ÖĞRENME TABANLI GÖRSEL ve ATALETSEL VERİ FÜZYONU

Mahmut KARAASLAN

Konya Teknik Üniversitesi
Lisansüstü Eğitim Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Doç. Dr. Ersin KAYA

2025, 57 Sayfa

Jüri
Doç. Dr. Ersin KAYA
Dr. Öğr. Üyesi Sedat KORKMAZ
Doç. Dr. Mehmet Akif ŞAHMAN

Son yıllarda İnsansız Hava Araçları (İHA) birçok uygulama alanında kullanılmaya başlanmıştır. İHA'ların kullanıldığı uygulamalarda konum bilgisinin doğruluğu bu uygulamaların başarısını doğrudan etkilemektedir. Konum bilgisi için yaygın olarak Küresel Konum Belirleme Sistemleri (Global Navigation Satellite System (GNSS)) kullanılmaktadır. Ancak GNSS kapalı veya zorlu ortamlarda kesintiye uğramaktadır. Bu durumun önüne geçmek ve sağlıklı bir konumlandırma sağlamak için GNSS sinyalinden bağımsız çalışabilen birçok yöntem önerilmiştir. Özellikle iç ortam konumlandırmada başarılı olan GNSS bağımsız yöntemlerde Görsel ve Ataletsel veriler ayrı ayrı veya birleştirilerek kullanılmıştır. Konumlandırma için yapılan ilk çalışmalarda klasik görüntü ve sensör işleme yöntemleri kullanılmaya başlansa da derin öğrenme alanında yaşanan gelişmeler bu alanı da etkilemiştir. Son yıllarda konumlandırmada klasik yöntemler ile yapılan birçok işlem için derin öğrenme tabanlı yöntemler önerilmiş ve başarılı olmuştur. Derin öğrenme tabanlı yöntemler klasik yöntemlerin önüne geçmeye başlasa da her iki yöntem türünün gelişimi devam etmiştir. Bu tez çalışmasında ise GNSS sinyal kayıplarının yaşandığı durumlarda İHA konumlandırılması amacıyla Derin Dikkat Tabanlı Görsel Ataletsel Konumlandırma Sistemi önerilmiştir. Görsel verilerin işlenmesi ve konumlandırma özelliklerinin çıkartılması için dikkat mekanizması entegre edilmiş Evrişimli Sinir Ağı (Convolutional Neural Network (CNN)) önerilmiştir. Ataletsel verilerin işlenmesi ve konumlandırma özelliklerinin çıkartılması için Dikkat Tabanlı Hiyerarşik Uzun Kısa Süreli Bellek (Attention-based Hierarchical Long Short-Term Memory (AHLSTM)) modeli kullanılmıştır. Dikkat mekanizmaları ile elde edilen görsel ve ataletsel özellikler, derin öğrenme tabanlı birleştirme yapılarıyla entegre edilerek yüksek doğrulukta konum tahmini gerçekleştirilmiştir. Görsel ve ataletsel verilerde dikkat mekanizmasının kullanılması konum belirleme açısından öne çıkan bölgelere odaklanmayı sağlayarak önemli bir performans artışı sağlamıştır. Ayrıca tez kapsamında yalnızca Ataletsel Ölçüm Birimi (IMU) verilerine dayanan özgün bir konumlandırma yaklaşımı da sunulmuştur. Bu yöntemde IMU verileri doğrudan konum bilgisi üretmek yerine uzaydaki konumsal değişimi modellemek amacıyla kullanılmıştır. Geliştirilen sistemler açık kaynaklı ve simülasyon ortamında oluşturulan veri setlerinde test edilmiştir. Sonuçlar dikkat mekanizmalarının görsel ve ataletsel verilerle birlikte kullanıldığında GNSS bağımsız güvenilir ve yüksek doğrulukta konumlandırma sağladığını göstermektedir.

Anahtar Kelimeler: Derin Öğrenme, Dikkat Mekanizması, İHA, Konumlandırma, Sensör Füzyonu

ABSTRACT

MS THESIS

DEEP LEARNING BASED VISUAL AND INERTIAL DATA FUSION FOR UAV LOCALİZATİON IN CASE OF GNSS SIGNAL LOSSES

Mahmut KARAASLAN

Konya Technical University
Institute of Graduate Studies
Department of Computer Engineering

Advisor: Assoc. Prof. Dr. Ersin KAYA

2025, 57 Pages

Jury
Assoc. Prof. Dr. Ersin KAYA
Assoc. Prof. Dr. Sedat KORKMAZ
Assoc. Prof. Dr. Mehmet Akif ŞAHMAN

In recent years, Unmanned Aerial Vehicles (UAVs) have begun to be used in many application areas. In applications where UAVs are used, the accuracy of location information directly affects the success of these applications. Global Navigation Satellite Systems (GNSS) are widely used for location information. However, GNSS is interrupted in closed or difficult environments. In order to prevent this situation and provide a healthy positioning, many methods that can work independently of the GNSS signal have been proposed. Visual and Inertial data have been used separately or combined in GNSS independent methods, which are especially successful in indoor positioning. Although classical image and sensor processing methods were first used for positioning, developments in the field of deep learning have also affected this field. In recent years, deep learning-based methods have been proposed and successful for many operations performed with classical methods in positioning. Although deep learning-based methods have begun to surpass classical methods, the development of both types of methods has continued. In this thesis, Deep Attention-Based Visual Inertial Positioning System has been proposed for UAV positioning in cases where GNSS signal losses occur. A Convolutional Neural Network (CNN) with an integrated attention mechanism is proposed for processing visual data and extracting positioning features. Attention-based Hierarchical Long Short-Term Memory (AHLSTM) model is used for processing inertial data and extracting positioning features. Visual and inertial features obtained with attention mechanisms are integrated with deep learning-based fusion structures to achieve high accuracy location estimation. Using the attention mechanism in visual and inertial data provides a significant performance increase by focusing on prominent regions in terms of location determination. In addition, an original location approach based solely on Inertial Measurement Unit (IMU) data is presented within the scope of the thesis. In this method, IMU data is used to model locational change in space instead of directly producing location information. The developed systems are open source and have been tested on datasets created in a simulation environment. The results show that attention mechanisms provide reliable and highly accurate location independent of GNSS when used with visual and inertial data.

Keywords: Attention Mechanism, Deep Learning, Localization, Sensor Fusion, UAV

ÖNSÖZ

Bu yüksek lisans tezi, lisansüstü eğitimim süresince edindiğim bilgi ve birikimi akademik bir çalışma çerçevesinde derlememe olanak sağlamıştır. Tez çalışmam süresince karşılaştığım bilimsel ve teknik sorunların çözümünde değerli katkılarını esirgemeyen, bilgi ve tecrübesiyle çalışmama yön veren danışmanım Doç. Dr. Ersin Kaya'ya en içten teşekkürlerimi sunarım.

Ayrıca, bu süreçte her zaman yanımda olan, manevi destekleriyle beni motive eden aileme ve çalışmamın çeşitli aşamalarında fikirleriyle katkı sunan tüm arkadaşlarıma teşekkür ederim.

Mahmut KARAASLAN
KONYA-2025

İÇİNDEKİLER

SİMGELER VE KISALTMALAR

Simgeler

x_t : t zamanındaki giriş
y_t : Çıkış
$W^{(l)}$: Ağırlık matrisi
$b^{(l)}$: Bias vektörü
U : Önceki gizli durum için öğrenilen ağırlık matrisleri
f : Aktivasyon fonksiyonu
f^{7x7} : 7x7 konvolüsyon filtresi
$\Delta\, x$: x Ekseninde Değişim
$\Delta\, y$: y Ekseninde Değişim
$\Delta\, z$: z Ekseninde Değişim
μ : Ortalama Değer
σ : Standart Sapma
$\sigma(x)$: Sigmoid aktivasyon fonksiyonu
h_t : Gizli durum
h_{t-1} : Önceki zaman adımındaki gizli durum
$\hat{h}_t$: Aday gizli durum
i_t : Giriş kapısı çıktısı
$\hat{c}_t$: Aday hücre durumu
c_t : Güncel hücre durumu
o_t : Çıkış kapısı
z_t : Güncelleme kapısı
r_t : Reset kapısı
e_t : t anındaki dikkat skoru
α_t : t anındaki normalize edilmiş dikkat ağırlığı
exp : Üstel fonksiyon

Kısaltmalar

ADAM : Adaptif Momentum
AHLSTM : Attention-based Hierarchical Long Short-Term Memory
BiLSTM : Bidirectional LSTM
BRIEF : Binary Robust Independent Elementary Features
CBAM : Convolutional Block Attention Module
CNN : Convolutional Neural Network
DNN : Deep Neural Network
DPVO : Deep Patch Visual Odometry
DROID : Differentiable Recurrent Optimization-Inspired Design
EKF : Extended Kalman Filter
GNSS : Global Navigation Satellite System
GPS : Global Positioning System
GRU : Gated Recurrent Unit
KF : Kalman Filter
LSD : Line Segment Detector
LSTM : Long Short-Term Memory

MAE	:	Mean Absolute Error
MAPE	:	Mean Absolute Percentage Error
MAV	:	Micro Aerial Vehicle
MSE	:	Mean Squared Error
RARNN	:	Recurrent Auto-Regressive Neural Network
RCNN	:	Recurrent Convolutional Neural Network
RMSE	:	Root Mean Squared Error
RNN	:	Recurrent Neural Network
R^2	:	Coefficient of determination
VO	:	Visual Odometry
VIO:	:	Visual-Inertial Odometri
VSLAM	:	Visual Simultaneous Localization and Mapping
IMU	:	Inertial Measurement Unit
İHA	:	İnsansız Hava Aracı
İKA	:	İnsansız Kara Aracı
UWB	:	Ultra-wideband

1. GİRİŞ

İlk kullanımına askeri alanda başlanan İnsansız Hava Araçları (İHA) birçok uygulama alanında kullanılmaya başlanmıştır (Couturier ve Akhloufi, 2021). Arama kurtarma çalışmaları (Tomic ve diğ., 2012), yük taşımacılığı (Hadi ve diğ., 2014), yangın tespiti (Sudhakar ve diğ., 2020), akıllı ulaşım sistemleri (Menouar ve diğ., 2017), tarımsal ilaçlama (Faiçal ve diğ., 2017), 5G iletişim sistemleri (Li ve diğ., 2018), haritalama (Goncalves ve Henriques, 2015) İHA uygulama alanlarına örnek olarak verilebilir.

İHA'lar örnek gösterilen uygulama alanları ve diğer uygulamalarda otonom ya da yarı otonom kullanılmaktadır. İHA'ların otonom ya da yarı otonom olarak kullanılabilmesi için doğru bir şekilde konumlandırma yapabilmesi gerekmektedir. İHA'larda konumlandırma için Atalet Ölçü Birimi (Inertial Measurement Unit (IMU)), kamera, lidar, radar, derinlik kameraları gibi sensörler kullanılmaktadır. Maliyet etkinliği ve kolay entegrasyon gibi avantajları dolayısıyla araştırmacılar IMU ve görüntü tabanlı algoritmalar üzerinde yoğunlaşmıştır (Gyagenda ve diğ., 2022).

IMU temel olarak eylemsizlik ivmesini ölçen ivmeölçer ve açısal dönüşü ölçen jiroskoptan oluşmaktadır (Ahmad ve diğ., 2013). Ataletsel sensörler önyargı, ölçek faktörü hatası ve kurulum hatası olmak üzere çeşitli deterministik hatalara sahiptir (Zheng ve diğ., 2023). Ataletsel sensörlerdeki hatalar zamanla birikerek Ataletsel Navigasyon Sisteminde (Inertial Navigation System (INS)) önemli sapmalara neden olur (Du ve diğ., 2017). Biriken hatalar ile INS'de oluşan sapmaların önüne geçmek için IMU ile birlikte tamamlayıcı olarak Küresel Konum Belirleme Sistemler (Global Navigation Satellite System (GNSS)) önerilmiştir (Wen ve diğ., 2021).

GNSS sinyalleri kapalı ortamlarda (Zafari ve diğ., 2019) veya şehirler gibi karmaşık ortamlarda kesintiye uğramaktadır (Bachrach ve diğ., 2011). Buna ek olarak GNSS sinyali siber saldırılar ile yanıltıcı sinyale dönüştürülebilmektedir (Kerns ve diğ., 2014). Bu yüzden GNSS sinyali olmadığı durumlarda veya yanlış çalıştığı durumlarda doğru konumlandırma yapabilecek algoritmaların geliştirilmesi gerekmektedir. GNSS tabanlı sistemler Kalman Filtresi (KF) gibi filtrelerle sensör füzyonu yaparak hassas konumlandırma yapmaktadır (Caron ve diğ., 2006).

Filtre tabanlı yöntemler uzun süreli sinyal kesintilerine karşı yeterince dayanıklı değildir. IMU verilerinde biriken hataları gidermek ve doğru konumlandırma sağlamak amacıyla, geleneksel filtre tabanlı yöntemlerin yanı sıra çoklu robot iş birliğine dayanan yöntemler de geliştirilmiştir. (Li ve diğ., 2022). Birden fazla İHA gereksinimi ve işbirlikçi

İHA'lar tarafından sağlanan bilgilerdeki düşük doğruluk işbirlikçi sistemlerin uygulanabilirliğini zorlaştırmaktadır. IMU'daki bilgilerden doğru çıktıların sağlanması için araştırmacılar öğrenme tabanlı algoritmalı kullanmaya yönelmişlerdir (Silva do Monte Lima ve diğ., 2019). IMU verilerinin derin öğrenme tabanlı yöntemler ile işlenmesi için zaman serisi modelleri kullanılmıştır (Jiang ve diğ., 2018).

Konumlandırma için kullanılan diğer bir sensör ise kameralardır. Kameralar ile konumlandırma için Eş Zamanlı Konum Belirleme ve Haritalama (Visual Simultaneous Localization and Mapping (VSLAM)) ve Görsel Odometri (Visual Odometry (VO)) yaygın olarak kullanılan iki görsel konumlandırma yöntemidir. (Bai ve diğ., 2023). VSLAM, bir robotun üzerindeki kamera yardımıyla hem kendi konumunu belirlemesini hem de çevresinin haritasını eş zamanlı olarak oluşturmasını sağlayan bir tekniktir. VO ise bir robotun hareketini ve konumunu ardışık görüntülerden tahmin eden bir tekniktir. VSLAM temelde VO tarafından elde edilen bilgileri kullanarak haritalama yapmaktadır (Gupta ve Fernando, 2022). Araştırmacılar, gerçek zamanlı konumlandırma gibi harita gerektirmeyen konularda hesaplama yükü fazla olan VSLAM yerine VO tabanlı algoritmaları tercih etmişlerdir (He ve diğ., 2020). VO temelde görüntü üzerindeki özellikleri yakalayarak hesaplama yapmaktadır. Ancak klasik görüntü özellik yakalama yöntemleri dokusuz ortamlardan olumsuz etkilenebilmektedir. Bu problemi çözmek için araştırmacılar derin öğrenme tabanlı algoritmalar yardımıyla özellik çıkarma ve konumlandırma yöntemlerine yönelmişlerdir (Chang ve diğ., 2023).

Çevresel etkileri en aza indirmek, ölçek belirsizliği problemini çözmek ve daha gürbüz (robust) bir konumlandırma sağlamak için VO ile IMU sensörünü birleştiren Görsel Ataletsel Odometri (Visual-Inertial Odometri (VIO)) sistemleri önerilmiştir (Zhuang ve diğ., 2024). İHA'larda konumlandırma için son gelişmeler hem İMU hem de kamera sensörünün yapay zekâ tabanlı olarak kullanılıp iki sensörün çıktılarının birleştirilmesi üzerinedir.

Bu tez çalışmasında görsel ve ataletsel verilerin derin dikkat tabanlı özellik çıkarımı ve füzyon yöntemiyle birleştirilmesine dayanan özgün bir konumlandırma sistemi önerilmiştir. Sistem mimarisinde IMU verilerinden zaman serisi modelleri ve görüntü verilerinden Evrişimli Sinir Ağı (Convolutional Neural Network (CNN)) modeli kullanılarak özellik çıkarımı gerçekleştirilmiştir. Elde edilen görsel-ataletsel özellikler derin öğrenme modelleri aracılığıyla birleştirilmiştir. Önceki çalışmalarda CNN modeli dikkat tabanlı mekanizmalar ile kullanıldığında konumlandırma için özellik seçiminde önemli iyileştirmeler sağlamıştır (Zhu ve diğ., 2022). Bu nedenle çalışmamızda dikkat

tabanlı bir görsel özellik çıkartma modeli kullanılmıştır. Ataletsel verilerin işlenmesinde de yapılan çalışmalarda GNSS kesintileri sırasında konumlandırma performansını önemli oranda artırmasından dolayı dikkat tabanlı zaman serisi modeli kullanılmıştır (Taghizadeh ve Safabakhsh, 2023). Önerilen Derin Dikkat Tabanlı Görsel Ataletsel Konumlandırma Sistemi CNN tabanlı bir özellik çıkarıcısı kullanıldığı için literatürdeki diğer yöntemlerden farklı olarak veri karıştırma (Shuffle) işlemi uygulanarak eğitim gerçekleştirilmiştir.

Tez kapsamında ayrıca sadece ataletsel verileri kullanan yenilikçi bir yöntem önerilmiştir. Önerilen sistemin literatürdeki çalışmalardan temel farkı IMU verilerinin işlenme yaklaşımında ortaya çıkmaktadır. Mevcut çalışmalardan farklı olarak İHA'nın IMU verisi doğrudan konum verisi olarak değil konumsal değişimin karakterizasyonu için kullanılmaktadır. Bu yaklaşım İnsansız Kara Araçlarında (İKA) başarıyla uygulanan IMU tabanlı eksensel değişim modellemesinin İHA sistemlerine adaptasyonunu temel almaktadır (Bai ve diğ., 2022).

Tez kapsamında önerilen yöntemler açık kaynak ve özgün olarak oluşturulan veri setlerinde test edilmiştir. Önerilen derin dikkat tabanlı görsel ataletsel konumlandırma yöntemi açık kaynak verilerde test edilmiş ve literatürdeki diğer çalışmalar ile karşılaştırılmıştır. Önerilen ataletsel konumlandırma sistemi ise bir simülasyon verisi kullanılarak test edilmiştir. Ataletsel konumlandırma sisteminde klasik ve zaman serisi tabanlı derin öğrenme modelleri ayrı ayrı denenerek sonuçları karşılaştırılmıştır. Ataletsel model veri setlerindeki eksik sensör verilerinden dolayı literatürdeki veri setleri ile kullanılamamış ve karşılaştırılamamıştır. Bu nedenle farklı konumlandırma verilerinin konumlandırmaya etkisini inceleyen deneysel bir çalışma yapılmıştır.

2. KAYNAK ARAŞTIRMASI

Bu bölümde literatürde bulunan temel konumlandırma çalışmaları ve İHA üzerine olan konumlandırma çalışmaları incelenmiştir. Görsel ve Ataletsel konumlandırma için birçok farklı çalışma yapılmıştır. Yapılan çalışmaların bir kısmı diğerlerinin eksikliklerini gidermek için güncelleme şeklinde bir kısmı ise yenilikçi yaklaşımlar olmuştur. Derin öğrenme alanında yaşanan gelişim bu alandaki çalışmaları etkilemiştir. Bununla birlikte araştırmacılar problemin türü, sensör maliyeti veya sistem gereksinimlerinin değişiklik göstermesi nedeniyle farklı temeldeki algoritmalardaki geliştirmelerini devam etmektedir.

Bölüm 2.1'de Klasik VO ve VIO alanında kullanılan klasik yöntemlerden bahsedilmiştir. Klasik yöntemler başlarda yalnız VO ile çalışmakta iken son zamanlarda VIO ile desteklenmiştir. Bölüm 2.2'de Öğrenme Tabanlı VO ve 2.3 Öğrenme Tabanlı VIO'da derin öğrenme tabanlı yöntemlerin konumlandırmadaki gelişimi incelenmiştir. Her iki alandaki sürekli gelişim ve yöntemsel farklılıklar nedeniyle, Öğrenme Tabanlı VO ve VIO ayrı başlıklar altında ele alınmıştır.

2.1. Klasik VO ve VIO

Mur-Artal ve diğ. (2015) özellik tabanlı monoküler SLAM uygulaması olan ORB-SLAM'i tanıtmışlardır (Mur-Artal ve diğ., 2015). Yönlendirilmiş HIZLI ve döndürülmüş BRIEF (Oriented FAST and rotated BRIEF (ORB)) özellik detektörü (Rublee ve diğ., 2011), diğer yöntemlere göre daha hızlı olmakla birlikte yüksek performans göstermesi sayesinde ORB-SLAM'in temelini oluşturmaktadır. ORB-SLAM, gerçek zamanlı, güvenilir, birden fazla senaryoda çalışabilen bir yöntem olarak birçok çalışmada kullanılmıştır. Sonraki yıllarda ise ORB-SLAM algoritmasına farklı problemlere hizmet eden yeni özellikler kazandırılmıştır. Mur-Artal ve Tardós (2017) ORB-SLAM2 ile monocular, stereo ve RGB-D kameralarda çalışabilen tam bir Görsel SLAM yöntemi tanıtmışlardır (Mur-Artal ve Tardós, 2017). ORB-SLAM'in son versiyonu olarak Campos ve diğ. (2021) ORB-SLAM3'ü tanıtmıştır (Campos ve diğ., 2021). ORB-SLAM3 algoritmasının en önemli yeniliklerinden biri görsel ataletsel SLAM sistemidir. Yapılan bu yenilikle ORB-SLAM algoritması görsel veriler ile ataletsel veriler birlikte kullanabilme özelliği kazanmıştır. Visual-inertial SLAM sistemi ile doğruluk önemli ölçüde artırılmış ve son derece gürbüz bir yaklaşım geliştirilmiştir.

Leutenegger ve diğ. (2015) OKVIS algoritmasını sunmuşlardır. OKVIS algoritması, görsel-ataletsel birleştirme yöntemlerinde sıkı bağlı bir yaklaşıma dayanan, anahtar kare (keyframe) tabanlı bir VIO sistemidir (Leutenegger ve diğ., 2015). Bu algoritma, ataletsel ölçümler ve görüntü anahtar noktalarının doğrusal olmayan optimizasyonu üzerine inşa edilmiş olup, hesaplama karmaşıklığını azaltırken yüksek doğruluğu korumaktadır. OKVIS, anahtar kare tabanlı yapı sayesinde sınırlı donanım kaynakları üzerinde gerçek zamanlı performans sağlamakta ve doğrusal olmayan optimizasyon sayesinde geleneksel doğrusal yöntemlerin ötesinde gelişmiş bir performans sunmaktadır. Bu algoritmanın son versiyonu Leutenegger (2022) tarafından tanıtılan OKVIS2'dir (Leutenegger, 2022). OKVIS2, uzun ve tekrar eden döngü kapanışlarını etkili bir şekilde yönetebilen yenilikçi bir Görsel Ataletsel SLAM sistemidir. Sistem, poz-graf kenarlarını ortak gözlemlerden marjinalleştirerek oluşturur ve döngü kapanışlarında bu kenarları tekrar konum noktalarına dönüştürür. Gerçek zamanlı optimizasyon yeteneği ile büyük ölçekli ortamlar için yüksek doğruluk sağlamakta ve asenkron olarak büyük döngüleri optimize edebilme kapasitesine sahiptir.

Qin ve diğ. (2018) sıkı sıkıya bağlı bir VIO olan VINS-Monoyu tanıtmışlardır (Qin ve diğ., 2018). VINS-Mono gürbüz bir başlatma, sıkı sıkıya bağlı bir VIO, online yeniden yerelleştirme gibi konumlandırma için önemli özelliklere sahiptir. Özellik detektörü ve tanımlayıcısı olarak İkili Gürbüz Bağımsız Temel Özellikler (Binary Robust Independent Elementary Features **(**BRIEF)) yöntemini kullanılmışlardır. Bulunan özelliklerin takibi için seyrek optik akış algoritmasından yararlanmışlardır. Doğrusal olmayan optimizasyon tabanlı bir yöntem ile IMU ile görsel özellikleri gürbüz bir şekilde birleştirmişlerdir. Çalışmada DBoW2 ile (Gálvez-López ve Tardos, 2012), döngü algılama sağlanmıştır. VINS-Mono'nun özellik tabanlı yapısının, ortam koşullarından olumsuz etkilenmesi nedeniyle geliştirilme ihtiyacı duyulmaktadır.

Rosinol ve diğ. (2020) semantik tabanlı konumlandırma ve haritalama yapan Kimera'yı tanıtmışlar (Rosinol ve diğ., 2020). Kimera ağ yeniden yapılandırmsası ve anlamsal etiketleme sayesinde ORB-SLAM, OKVIS, VINS-Mono gibi algoritmaların önüne geçmiştir. Kimera Georgia Tech Düzeltme ve Haritalama Kütüphanesi (Georgia Tech Smoothing and Mapping Library (GTSAM)) üzerine kurulmuş bir VIO algoritmasına sahiptir (Dellaert, 2012). Kimera diğer SLAM yöntemlerinde çok daha farklı ve yenilikçi olarak semantik segmentasyondan yararlanmaktadır. Kimera gerçek zamanlı ve metrik semantik SLAM bir uygulama olarak konumlandırma alanında önemli bir yenilik sunmaktadır.

Klasik yöntemlerde yapılan son çalışmalarda konumlandırma algoritmalarının çalışma sürelerinin hızlandırılması üzerine yoğunlaşılmıştır. Fu ve diğ. (2020) PL-VINS algoritmasını önermiştir (Fu ve diğ., 2020). PL-VINS nokta ve çizgi özelliklerini kullanan VINS-Mono üzerine geliştirilen bir algoritmadır. Çalışmada çizgi özelliklerini çıkarmak için kullanılan Çizgi Segment Dedektörü (Line Segment Detector (LSD)) algoritması (Von Gioi ve diğ., 2008), üzerinden bazı değişikliler yapılarak üç kat hız artışı sağlanmıştır. Hızlandırma üzerine yapılan bir diğer çalışma ise Seiskari ve diğ. (2022) tarafından sunulan HybVIO'dur (Seiskari ve diğ., 2022). VIO'nun SLAM ile birleştirildiği yeni bir hibrit yöntemdir. Çalışmada probolistik VIO geliştirilmiştir. HybVIO doğruluk değeri diğer çalışmalara yakın önemli bir gerçek zamanlı konumlandırma sistemi olarak öne çıkmaktadır.

2.2. Öğrenme Tabanlı VO

Kendall ve diğ. (2015) sadece görüntüler üzerinden CNN tabanlı konumlandırma yaklaşımı olan PoseNet'i sunmuşlardır (Kendall ve diğ., 2015). PoseNet CNN tabanlı uçtan uca eğitilebilir 6-DOF lokalizasyon yapabilen bir sistemdir. Wang ve diğ. (2017) Yinelemeli CNN (Recurrent CNN(RCNNs)) tabanlı VO olan DeepVo'yu tanıtmıştır (Wang ve diğ., 2017). DeepVo bir CNN mimarisi ile çıkardığı özellikleri Yinelemeli Sinir Ağı (Recurrent Neural Network (RNN)) ile birlikte kullanarak sıralı öğrenme sağlamıştır. Önerilen RCNN modeli ile VO için uçtan uca öğrenebilen bir yöntem sağlanmıştır. DeepVo sahip olduğu CNN tabanlı özellik temsili ile algoritma klasik yöntemlere göre dokusuz veya karmaşık ortamlarda klasik yöntemlere göre önemli bir potansiyele sahiptir. Araştırmacılar DeepVO'nun geometri tabanlı yöntemlerin yerine geçmesinden ziyade bu yöntemlere katkı sağlayan bir tamamlayıcı nitelikte olduğunu vurgulamışlardır. Valada ve diğ. (2018) VLocNet'i önermişlerdir (Valada ve diğ., 2018). VLocNet temelde Derin CNN mimarisine sahiptir. VLocnet küresel poz regresyonu ve bağıl poz tahmini birleşimiyle oluşmaktadır. Önerilen mimari 6 serbestlik derecesine (DoF) sahip konumlandırma ve odometri yapabilme yeteneğiyle mevcut CNN tabanlı yöntemlere kıyasla üstün bir performans sergilemiştir. Zhu ve diğ. (2022) hareket desenleri üzerine odaklanan DeepAVO'yu önermişlerdir (Zhu ve diğ., 2022). DeepAVO optik akış girdilerini kullanarak Dikkat tabanlı CNN aracılığıyla rotasyon ve çevirme bilgilerini etkili bir şekilde öğrenebilen bir mimaridir. Geliştirilen mimari, monoküler yöntemlere

göre üstün performans göstermiş olup, araştırmacılar tarafından gelecekte kapsamlı bir SLAM sistemine dönüştürülmesi hedeflenmektedir.

CNN ve derin öğrenme tabanlı tekniklerin genelleme kabiliyetleri eğitim aşamasında karşılaşmadıkları ortamlarda sınırlı kalmakta ve sistem performansında kayda değer bozulma gözlemlenmektedir. Bundan dolayı bazı araştırmacılar genellenebilir yöntemleri geliştirmeye odaklanmıştır (Li ve diğ., 2021). Teed ve Deng (2021) güçlü bir genelleme yeteneğine sahip DROID-SLAM'i tanıtmışlardır (Teed ve Deng, 2021). DROID-SLAM sadece sentetik veri üzerinde eğitilmiş olup gerçek dünya koşullarında toplanan farklı karakteristikteki veri setlerinde de etkin performans sergilemiştir. Ayrıca eğitim için yalnızca monoküler görüntüler kullanılmış ve bu model hem stereo hem de RGB-D kameralarda başarılı bir şekilde kullanılmıştır. Farklılaştırılabilir Tekrarlayan Optimizasyondan Esinlenen Tasarım (Differentiable Recurrent Optimization-Inspired Design (DROID)) klasik ve derin öğrenme tabanlı yöntemleri birbiri ile tamamlayıcı olarak kullanan optik akış tabanlı bir yöntemdir. Teed ve diğ. (2024) daha sonra Derin Yama Görsel Odometrisi (Deep Patch Visual Odometry (DPVO))'yu önermişlerdir (Teed ve diğ., 2024). DPVO video üzerinden kamera hareketlerini takip eden yineleyen ağdan oluşmaktadır. Yama tabanlı eşleşme ve demet ayarlaması için önerilmiş yeni bir yineleyen güncelleme operatörü tanıtılmıştır. DPVO DROID'e göre yaklaşık üç kat daha hızlı çalışmaktadır. DPVO sadece odometri sistemi barındırmaktadır. Bu yüzden DROID-SLAM'e göre dış ortamlarda başarısız olabilmektedir. Lipson ve diğ. (2024) DPVO algoritmasının eksikliklerini gidermek ve daha gürbüz konumlandırma yapmak için DPV-SLAM'i önermişlerdir (Lipson ve diğ., 2024). DPVO algoritmasına SLAM özelliğinin kazandırılması ile daha genellenebilir ve gürbüz bir yöntem oluşmuştur. DPV-SLAM sunduğu yenilikçi yaklaşımlarla birlikte üstün hesaplama performansı ve yüksek verimlilik göstermektedir.

2.3. Öğrenme Tabanlı VIO

Clark ve diğ. (2017) bilinen ilk uçtan uca eğitilebilir VIO tabanlı konumlandırma sistemi olan VINet'i tanıtmışlardır (Clark ve diğ., 2017). Önerilen mimaride CNN ile görüntü üzerinden LSTM ile IMU üzerinden özellikleri çıkartılarak bir araya getirilmiş ve odometri tahmini sağlanmıştır. Daha sonraki çalışmalarda Baldini ve diğ. (2020) VINet ile benzer bir yapıya sahip bir yöntem önermişlerdir (Baldini ve diğ., 2020). Önerdikleri yöntemde VINet'den farklı olarak ardışık iki frame yerine IMU ile tek bir

çerçeveyi eğitmişlerdir. Solodar ve Klein (2024) VIO-DualProNet'i tanıtmışlardır (Solodar ve Klein, 2024). Önerdikleri yöntem ile VINS-Mono'ya IMU kaynaklı gerçek zamanlı ataletsel gürültü belirsizliğini tahmin eden bir derin öğrenme modeli eklemişlerdir. Geliştirdikleri yöntem ile VINS-mono'da iyileştirme yapmışlar.

Aslan ve diğ. (2022) IMU ve Görsel veriyi CNN modelinde işleyip daha sonra birleştiren VIIONet'i önermişlerdir (Aslan ve diğ., 2022). VIIONet ataletsel ve görsel verileri Inception V3 modeli ile eğitip daha sonra bu modellerden gelen özellikleri birleştirerek konumlandırma yapmaktadır. IMU verisi Savitzky-Golay tekniği ile filtrelenip normalizasyon yapılır. Daha sonra elde edilen sayısal veriler CNN modeline verilmek için görüntüye dönüştürülür. Görsel veri ise ardışık çerçevelerde optik akış çıktıları alınıp birleştirildikten sonra modele verilmektedir. Daha sonra ataletsel ve görsel özellikler Gauss süreci regresyonu ile birleştirilerek konumlandırma çıktısı sağlanmaktadır (Rasmussen, 2003). VIIONet konumlandırmada oldukça yüksek doğruluk elde etse de yüksek işlemci gereksinimi ve gecikme süresi yöntemin gerçek zamanlı olarak çalışabilirliğini kısıtlamaktadır.

Aslan ve diğ. (2022) İHA'larda konum tahmini için yeni bir VIO sistemi olan Hibrit Görsel Eylemsiz Odometri (Hybrid Visual Inertial Odometry (HVIOnet)) yöntemini önermişlerdir. HVIOnet konumlandırma yapmak için derin öğrenme tabanlı kamera ve IMU sensör füzyonu içeren bir sistemdir. Derin öğrenme tabanlı füzyon işlemi için her iki sensör türü özellik çıkartıcı olarak kullanılmıştır. IMU üzerinden özellik çıkarımı için BiLSTM mimarisi sensör girdilerine karşılık 3D pozisyon verileri ile eğitilmiştir. Görüntü özelliklerinin çıkartılması için ise kamera görüntüleri CNN modeline verilerek konumlandırma çıktıları ile eğitilmiştir. Daha sonra her iki sensör türünden elde edilen ataletsel ve görsel özellikler BiLSTM sinir ağına verilerek tam bir konumlandırma çıktısı sağlanmıştır. Çalışma konumlandırma üzerine yapılmış diğer yöntemlerle karşılaştırıldığında dikkate değer sonuçlar elde etmiştir (Aslan, ve diğ., 2022).

Yang ve diğ. (2022), derin öğrenme tabanlı yöntemlerin yüksek maliyetle çalışmasından dolayı Visual Selective VIO yöntemini önermişlerdir. Bu çalışmada görüntü verilerinin IMU'ya kıyasla daha yüksek maliyetle çalışmasına rağmen sağladığı bilgilerin konumlandırma doğruluğunu aynı oranda artırmadığına değinilmektedir. Buna çözüm mevcut duruma göre görüntü özellik çıkartıcısını devre dışı bırakan adaptif bir yöntem önerilmiştir. Bu yöntemde uçtan uca öğrenmeyi sağlamak için Gumbel Softmax

benimsenmiştir. Önerilen yöntem belirgin bir performans kaybı olmadan önemli oranda hesaplama maliyetini düşürmüştür (Yang ve diğ., 2022).

Tu ve diğ. (2022), dışsal dikkat tabanlı EMA-VIO'yu önermişlerdir. Diğer derin öğrenme tabanlı yöntemlerde kullanılan RNN tabanlı konumlandırma modellerinin maliyeti olduğunu savunmuşlardır. Bunun için RNN tabanlı modeller yerine dışsal bir dikkat modeli ile sensör birleştirme işlemini yapmışlardır. Önerilen EMA-VIO mimarisi ile geleneksel yöntemlerin özellik çıkarmada zorlandığı kapalı havalar veya suyla kaplı zeminlerde başarılı bir şekilde konumlandırma sağlamışlardır (Tu ve diğ., 2022).

Kao ve diğ. (2023) Görüntü, IMU ve Ultra Geniş Band (Ultra-wideband (UWB)9 sensörlerini derin öğrenme ile birleştirerek konumlandırma yapan VIUNet'i önermişlerdir. VIUNet görüntü ile IMU verilerini işleyen bir dal ve UWB sensörlerini işleyen bir dal olmak üzere iki daldan oluşmaktadır. Görüntü verilerini işlemek için CNN tabanlı bir mimari ile optik akış hesaplaması yapılıp görsel özellik çıktısı sağlanmaktadır. IMU verilerini işlemek için LSTM mimarisi kullanılarak ataletsel özellik çıktısı sağlanmaktadır. Daha sonra bu özellikler ve UWB sensör bilgileri klasik bir derin sinir ağı ile birleştirilerek konumlandırma çıktısı sağlanmaktadır. UWB sensörünün önerilen yöntemde önemli oranda performans artışı sağladığını ortaya koymuşlardır (Kao ve diğ., 2023).

Pan ve diğ. (2024), geleneksel doğrusal olmayan (non-linear) optimizasyonu sürekli öğrenme ile birleştiren Adaptive VIO'yu önermiştir. Bu yöntemde literatürdeki diğer çalışmalardaki gibi görsel ve ataletsel verileri birleştirerek doğrudan tahmin yapmak yerine doğrusal olmayan bir algoritma olan demet ayarlaması ile birleştirerek konumlandırma çıktısı sağlamışlardır. Yapılan deneysel çalışmalar Adaptive VIO'nun açık kaynak veri setlerinde başarılı performans sergilediğini ve mevcut VIO tabanlı yöntemlerin önüne geçtiğini göstermektedir. Ayrıca güncel optimizasyon algoritmalarıyla rekabet edebilecek düzeyde olduğu da kanıtlanmıştır (Pan ve diğ., 2024).

Yuan ve diğ. (2024), geleneksel özellik çıkarma yöntemleri ile derin öğrenme tabanlı özellik çıkarma ve izleme yöntemlerini birleştiren Mix-VIO'yu önermiştir. Önerdikleri yöntemde derin öğrenme tabanlı yöntemlerinin hızını TensorRT kullanarak artırmışlardır. Derin öğrenme ve geleneksel optik akış yöntemlerini bir arada kullanarak hızlı kamera değişimleri ve değişken aydınlatma koşullarında başarılı sonuçlar elde etmişlerdir. Açık kaynak veri setlerinde yapılan testler ile doğruluk ve dayanıklılık açısından üstün performans sağladığı gösterilmiştir (Yuan ve diğ., 2024).

3. MATERYAL VE YÖNTEM

3.1. Veri Setleri

Tez kapsamında iki farklı yöntem önerilmiştir. Önerilen iki farklı yöntemi test etmek için iki farklı veri seti kullanılmıştır. Ataletse konumlandırma testi için 3.1.1.'de verilen Webots simülasyon ortamı kullanılarak özgün bir veri seti oluşturulmuştur. Oluşturulan özgün veri seti ile önerilen yöntem test edilmiştir. Derin Dikkat Tabanlı Görsel Ataletsel Konumlandırma sisteminin testi ve literatür karşılaştırması için 3.1.2.' de açıklanan açık kaynak veri seti kullanılmıştır.

3.1.1. Webots simülasyon ortamı veri seti

Webots simülasyon ortamı, robotik simülasyon ve geliştirme için kullanılan bir platformdur. Bu simülasyon ortamı, kullanıcıların robotların hareketlerini ve davranışlarını sanal bir ortamda test etmelerine olanak tanımaktadır. Webots, fiziksel motoruyla gerçekçi simülasyonlar sağlar ve robotların sensörlerini, aktüatörlerini ve çevresel etkileşimlerini detaylı bir şekilde modelleyebilir. Kullanıcılar, C, C++, Python, Java ve MATLAB gibi çeşitli programlama dilleriyle robotları programlayabilir ve bu programları simülasyon ortamında çalıştırabilirler. Ayrıca Webots, 3D modelleme yetenekleri sayesinde kullanıcıların özel robot tasarımlarını kolayca oluşturup test etmelerine imkân tanır, bu da gerçek dünya uygulamalarına kıyasla maliyet ve zaman tasarrufu sağlar. Webots simülasyon ortamının farklı programlama dillerini desteklemesi, robot davranışlarının programlanabilmesi ve hazır bileşenlerin ortama aktarılması gibi avantajlara sahip olmasından dolayı birçok araştırmacı tarafından kullanılmıştır (Michel, 2004).

Veri seti toplama işlemi için webots simülasyon ortamında bir çevre tasarlanmıştır. Bu çevre içerisinde dji mavic2 pro marka İHA yerleştirilmiştir. Simülasyondaki İHA üzerine İMU, jiroskop, ivmeölçer, pusula ve GPS sensörleri yerleştirilmiştir. Bu çalışma için simülasyon ortamında 9 farklı uçuştan oluşan toplamda 8 dakika 20 saniyelik veri seti oluşturulmuştur.

3.1.2. European robotics challenge micro aerial vehicle (EuRoc MAV) veri seti

EuRoc MAV veri seti (Burri ve diğ., 2016), ETH Zurich tarafından geliştirilmiş ve SLAM araştırmalarında yaygın olarak kullanılan bir veri setidir. Veri seti AscTec Fireflt MAV aracı görsel ve ataletsel sensörler donatılarak toplanmıştır. Görüntü verisini sağlamak için gri renk görüntüsü ve 20 HZ hıza sahip 2 adet MT9V034 kamera kullanılmıştır. Atalet verisini kaydetmek için 200 HZ hıza sahip jiroskop ve ivmeölçerden oluşan ADIS16448 model IMU kullanılmıştır. Gerçek konumlandırma verilerini oluşturmak için 20 HZ hıza sahip Leica Nova MS50 model lazer kullanılmıştır. Veri setinin çeşitliliğini arıtmak için aynı donanımla farklı ortam ve zorluklarda uçuşlar yapılmıştır. Bu sayede yöntemlerin farklı koşullarda test edilmesi ve karşılaştırılması mümkün olmuştur.

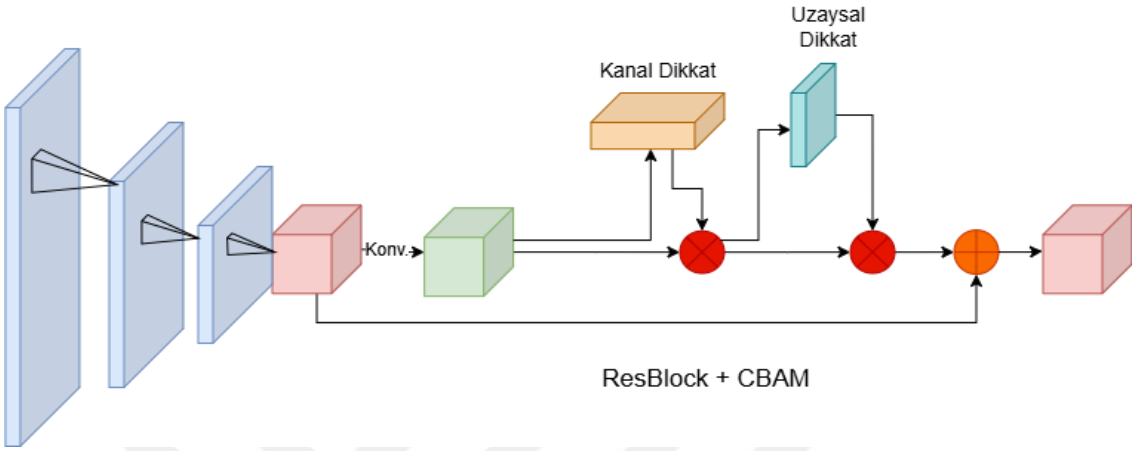
Bu veri setinde Vicon Room 1 ve Vicon Room 2 olmak üzere iki farklı ortam vardır. Her ortam için kolay, orta, zor olarak kategorilendirilen veri toplanmıştır. Kolaylık derecesi ortam karmaşıklığı ve agresif manevralar gibi etmenlere göre belirlenmiştir. “V1” ve “V2” farklı odaları “01”, “02”, “03” ise sırasıyla kolay, orta, zor kategorilerini temsil edecek şekilde V1-01, V1-02, V1-03, V2-01, V2-02 ve V2-03 olmak üzere 6 farklı Vicon Room veri seti kullanılmıştır.

3.2. Derin Öğrenme Modelleri

3.2.1. Dikkat tabanlı CNN modeli

Görsel özelliklerin çıkartılması için dikkat tabanlı bir CNN mimarisi kullanılmıştır. CNN mimarilerine dikkat mekanizmasını uygulayabilmek için Konvolüsyonel Blok Dikkat Modülü (Convolutional Block Attention Module (CBAM)) tanıtılmıştır. Konvolüsyon işlemlerinde kanal (Channel) ve uzay (Spatial) bilgileri bir araya getirerek özellik çıkarımı yapılmaktadır. CBAM algoritmasında bunu yapmak için Kanal Dikkat (Channel Attention) ve Uzaysal Dikkat (Spatial Attention) modülleri geliştirilmiştir. Bu modüller sayesinde CNN modelinde Kanal boyutunda “ne” ve uzaysal boyutta “nerede” dikkat edeceğini öğrenebilecektir. Geliştirilen iki boyutta çalışan modüllere girdi özellikleri verilerek elde dikkat haritaları elde edilir. Elde edilen dikkat haritaları girdi özellikleri ile çarpılarak iyileştirilmiş özellikler çıktı olarak sağlanır (Woo ve diğ., 2018).

CBAM genel bir modül olarak tasarlanmış olup herhangi bir CNN mimarisine entegre edilebilmektedir. Çalışmamızda CNN mimarisi olarak ResNet50 kullanılmıştır. ResNet50 modelinin son katmanına CBAM eklenerek hedeflenen dikkat mekanizması sağlanmıştır. Önerilen dikkat tabanlı CNN modeli Şekil 3.1' de verilmiştir.



Şekil 3.1. Dikkat Tabanlı CNN Modeli

Kanal dikkat modülü, bir giriş özellik haritası F üzerinde hangi kanalların daha önemli olduğunu belirlemeyi amaçlar. Bu amaçla, önce F üzerinde ortalama havuzlama (Average Pooling) ve maksimum havuzlama (Max Pooling) işlemleri ayrı ayrı uygulanır. Bu işlemler, özellik haritasının her kanalını bir skaler değere indirger. Daha sonra elde edilen bu özet temsiller paylaşımlı ağırlıklara sahip iki katmanlı birçok katmanlı algılayıcı (MLP) ile işlenir. MLP'nin çıktıları toplanarak birleştirilir ve ardından sigmoid aktivasyon fonksiyonu (σ) ile normalize edilir. Böylece her kanal için 0 ile 1 arasında bir dikkat skoru elde edilir. Elde edilen bu kanal dikkat haritası ($M\_c(F)$), giriş özellik haritası (F) ile eleman bazlı olarak çarpılır ve kanal boyutunda dikkat uygulanmış olur. Kanal dikkat modülü Denklem (3.1) ile ifade edilir.

$$M\_c(F) = \sigma(MLP(AvgPool(F)) + MLP(MaxPool(F))) \quad (3.1)$$

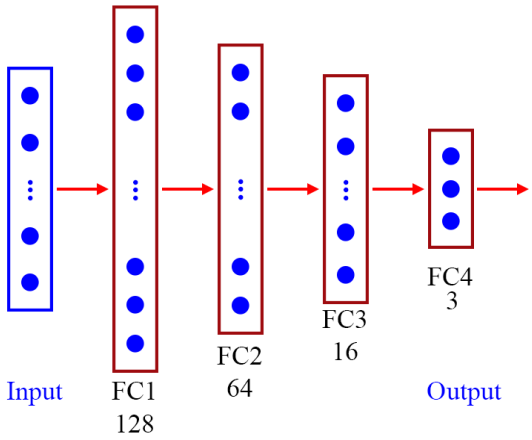
Uzaysal dikkat modülü, kanal dikkatinden elde edilen çıktı olan F' üzerinden çalışır. Bu modülde, önce F' üzerinde kanal boyunca ortalama havuzlama (Average Pooling) ve maksimum havuzlama (Max Pooling) yapılır. Böylece her bir uzamsal konumun ortalama ve maksimum değerleri elde edilir. Bu iki harita daha sonra birleştirilerek bir giriş olarak kullanılır ve bir adet 7×7 boyutlu konvolüsyon filtresi

(f^{7x7}) üzerinden geçirilir. Elde edilen sonuç, sigmoid aktivasyon fonksiyonu (σ) ile normalize edilerek uzaysal dikkat haritası ($M\_s(F')$) elde edilir. Bu dikkat haritası, giriş haritası (F') ile çarpılarak her konum için dikkat uygulanmış olur. Uzaysal dikkat modülü Denklem (3.2) ile ifade edilir.

$$M\_s(F') = \sigma(f^{7x7}([AvgPool(F'); MaxPool(F')])) \quad (3.2)$$

3.2.2. Derin sinir ağı (Deep neural network (DNN))

DNN'ler, girdi verileri arasındaki karmaşık, hiyerarşik ve non-lineer ilişkileri öğrenme yeteneği ile bilinir. İHA'ların konum tahmini için kullanılan sensör verileri genellikle bu karmaşıklığı içerir. Örneğin, ivmeölçer, jiroskop, gibi sensörlerden alınan verilerin bir araya getirilerek konum tahminine dönüştürülmesinde DNN'ler etkili olabilir. Bu yüzden çalışma kapsamında DNN modelleri ile eğitim işlemi gerçekleştirilmiştir. Tasarlanan örnek bir DNN modeli Şekil 3.2'de verilmiştir.



Şekil 3.2. Derin Sinir Ağı Modeli

Derin sinir ağı (DNN), girişten çıkışa kadar birden çok tam bağlantılı (fully connected, FC) katmandan oluşur. Her bir katman, bir önceki katmandan gelen çıktıyı işleyerek bir sonraki katmana aktarır. l. katmandaki çıkış aktivasyonu ($h^{(l)}$), o katmana ait ağırlık matrisi ($W^{(l)}$), bias vektörü ($b^{(l)}$), ve bir aktivasyon fonksiyonu (f) kullanılarak hesaplanır. Burada $h^{(l-1)}$, bir önceki katmanın çıktısını temsil eder. Hesaplama, ağırlıklı toplam ve bias eklenmesinden sonra aktivasyon fonksiyonunun uygulanmasıyla gerçekleştirilir. Bu işlem, tüm katmanlar boyunca tekrarlanarak ağın

girişten çıkışa kadar öğrenme ve temsil kapasitesi artırılır. Bu yapı sayesinde DNN, karmaşık ve doğrusal olmayan ilişkileri öğrenebilmektedir. DNN yapısı Denklem (3.3) ile ifade edilebilmektedir.

$$h^{(l)} = f(W^{(l)} h^{(l-1)} + b^{(l)}) \quad (3.3)$$

3.2.3. Yinelmeli sinir ağı (Recurrent neural network (RNN))

İHA'ların sensör verileri genellikle zamanla değişen bir yapı gösterir. RNN, önceki zaman adımlarındaki bilgileri kullanma yeteneği ile zaman serisi verilerini etkili bir şekilde işleyebilir. Bu yüzden çalışma kapsamında RNN modellerinin kullanılmasına karar verilmiştir.

RNN'ler özellikle sıralı veriler üzerinde çalışan ve zaman içindeki bağımlılıkları yakalamak için tasarlanmış bir yapay sinir ağı türüdür. Geleneksel yapay sinir ağlarından farklı olarak RNN'ler döngüsel bağlantılara sahiptir. Yani önceki zaman adımlarındaki bilgiyi mevcut zaman adımına taşıyabilirler. Bu özellik metin, ses veya zaman serisi gibi verilerdeki örüntüleri ve ilişkileri anlamada çok etkilidir. RNN'ler dil modelleme, çeviri, konuşma tanıma ve zaman serisi tahmini gibi uygulamalarda yaygın olarak kullanılır. Ancak kaybolan ve patlayan gradyanlar problemlerine sahiptir. Bu problemlerin önüne geçmek için çeşitli geliştirilmiş versiyonları mevcuttur.

RNN'ler zaman bağımlı verilerdeki korelasyonları Denklem (3.4) ve Denklem (3.5) ile modellemektedir.

$$h_t = tanh(W_{ih} x_t + W_{hh} h_{t-1} + b_h) \quad (3.4)$$

$$y_t = W_{ho} h_t + b_o \quad (3.5)$$

Yukarıdaki formülde, x_t ifadesi t zaman adımındaki giriş verisini, h_t ise bu zamandaki gizli durumu temsil etmektedir. Gizli durum, girişten gelen bilgi ile bir önceki gizli durum olan h_{t-1} 'in lineer kombinasyonunun, tanjant hiperbolik ($tanh$) aktivasyon fonksiyonuyla işlenmesiyle elde edilir. Bu işlem sırasında W_{ih} girişten gizli katmana olan ağırlık matrisini, W_{hh} gizli katmanın kendi içindeki ağırlıklarını ve b_h da bias (sapma) terimini ifade eder. Modelin çıktısı olan y_t, güncel gizli durum h_t 'nin W_{ho} ağırlık matrisiyle çarpılması ve $b_o$bias terimi ile toplanmasıyla elde edilir. Bu yapı, modelin zaman içindeki bağlamı öğrenmesini sağlar.

3.2.4. Uzun ömürlü kısa-dönem belleği (Long short-term memory (LSTM))

Long Short-Term Memory (LSTM) (Hochreiter ve Schmidhuber, 1997), RNN'lerin uzun vadeli bağımlılıkları daha etkili bir şekilde öğrenmesini sağlamak için tasarlanmış özel bir türüdür. LSTM'ler, bilgi akışını kontrol eden giriş, unutma ve çıkış kapıları gibi bileşenlere sahiptir. Bu kapılar, hangi bilginin hücreye girip çıkacağını ve hangi bilginin tutulacağını veya unutulacağını belirler. Bu sayede LSTM'ler, uzun sekanslardaki bağımlılıkları daha iyi öğrenebilir ve dil modelleme, çeviri, konuşma tanıma ve zaman serisi tahmini gibi uygulamalarda daha başarılı sonuçlar elde eder. LSTM'ler, özellikle geleneksel RNN'lerin başa çıkmakta zorlandığı uzun vadeli ilişkileri yakalamada oldukça etkilidir.

LSTM yapısındaki unutma kapısı, önceki hücre durumunun hangi kısmının unutulacağına karar verir. Eğer f_t değeri 0'a yakınsa ilgili bilgi unutulur, 1'e yakınsa korunur. Bu kapının çıktısı, sigmoid aktivasyon fonksiyonu ile hesaplanır ve Denklem (3.6) ile ifade edilir.

$$f_t = \sigma(Wf\, x_t + Uf\, h_{t-1} + bf) \tag{3.6}$$

Burada x_t t anındaki giriş verisini, h_{t-1} ise bir önceki zaman adımındaki gizli durumu temsil eder. Wf, Uf ve bf ise sırasıyla giriş, gizli durum ve bias için öğrenilen parametrelerdir.

Ardından, giriş kapısı yeni bilginin hücreye ne kadar aktarılacağını belirler. Giriş kapısının çıktısı Denklem (3.7) ile hesaplanır. Aynı zamanda aday hücre durumu, yani güncellenmeye hazır yeni bilgi ise Denklem (3.8) ile hesaplanır.

$$i_t = \sigma(Wi\, x_t + Ui\, h_{t-1} + bi) \tag{3.7}$$

$$\hat{c}_t = tanh(Wc\, x_t + Uc\, h_{t-1} + bc) \tag{3.8}$$

Yeni hücre durumu c_t unutma kapısı tarafından belirlenen ölçüde önceki hücre durumu c_{t-1}'in korunması ve giriş kapısı aracılığıyla aday bilginin eklenmesiyle güncellenir.

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \hat{c}_t \tag{3.9}$$

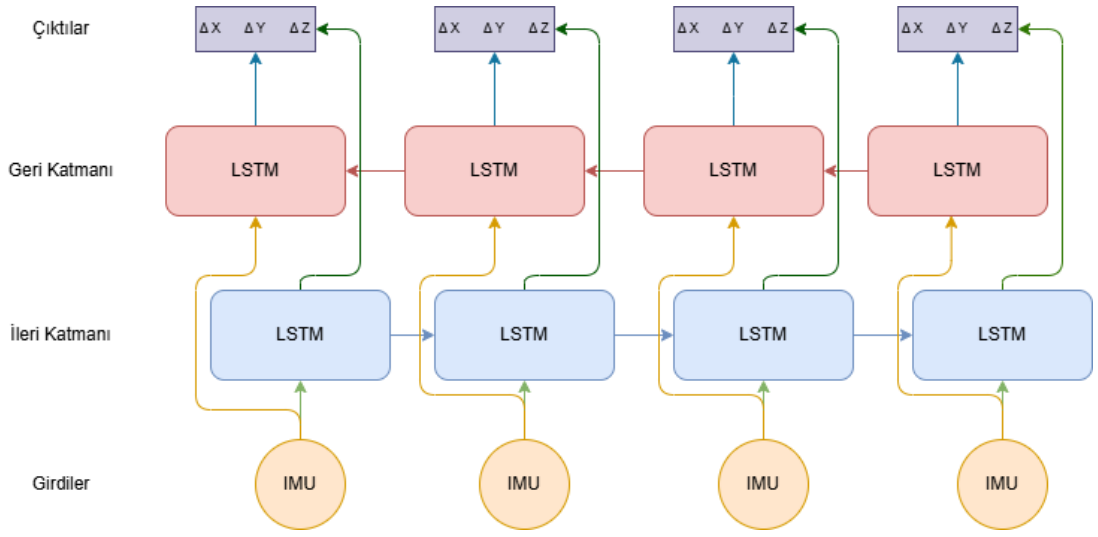
Son olarak, çıkış kapısı hücre durumunun ne kadarının çıktıya dönüştürüleceğini Denklem (3.10)'deki ifade ile belirler. Hücre durumu önce tanh ile yumuşatılır ve çıkış kapısı tarafından ağırlıklandırılarak dışarıya aktarılır (Denklem (3.11)). Bu yapı sayesinde, hücre durumu (c_t) zaman içinde gerekli olan bilgileri koruyarak uzun vadeli bağımlılıkları güçlü bir şekilde temsil edebilmektedir.

$$o_t = \sigma(Wo\, x_t + Uo\, h_{t-1} + bo) \quad (3.10)$$

$$h_t = o_t \cdot tanh(c_t) \quad (3.11)$$

3.2.5. Çift yönlü LSTM (Bidirectional LSTM (BiLSTM)

Bidirectional Long Short-Term Memory (BiLSTM), sekans verilerinde hem geçmiş hem de gelecekteki bağlamı öğrenmek için tasarlanmış bir LSTM varyantıdır. BiLSTM, veriyi iki yönlü olarak işler: birinci yön, veriyi zamanın ileri akışında işlerken, ikinci yön, aynı veriyi zamanın geri akışında işler. Bu iki yönlü bilgi akışı her bir zaman adımında hem önceki hem de sonraki bağlamı yakalayarak daha zengin ve kapsamlı bir özellik seti oluşturur. Bu özellik özellikle dil işleme, konuşma tanıma ve metin sıralama gibi uygulamalarda kelimelerin hem önceki hem de sonraki kelimelerle olan ilişkilerini daha iyi anlamayı sağlar. BiLSTM'ler, LSTM'lerin tüm avantajlarını korurken, iki yönlü bilgi akışı sayesinde daha güçlü ve esnek bir model sunar. BiLSTM yapısı şekil 3.3'te verilmiştir.



Şekil 3.3. BiLSTM Yapısı

BiLSTM, LSTM'nin hem geçmiş hem de gelecek zaman adımlarını hesaba katan bir versiyonudur. İki adet LSTM paralel olarak çalışır (Denklem (3.12)-Denklem (3.14)).

$$\rightarrow h_t = LSTMfwd(x_t, \rightarrow h_{t-1}) \quad (3.12)$$

$$\leftarrow h_t = LSTMbwd(x_t, \leftarrow h_{t+1}) \quad (3.13)$$

$$h_t = [\rightarrow h_t; \leftarrow h_t] \quad (3.14)$$

Burada x_t t anındaki giriş verisini, $\rightarrow h_t$ t anındaki ileri yönlü LSTM gizli durumunu (geçmişe bakar), $\leftarrow h_t$ t anındaki geri yönlü LSTM gizli durumunu (geleceğe bakar) ifade eder. $LSTMfwd$ ve $LSTMbwd$ ise sırasıyla ileri yönlü LSTM ve geri yönlü LSTM hücresini temsil eder. h_t ise t zaman adımındaki BiLSTM çıktısıdır

3.2.6. Kapılı yinelemeli üniteler (Gated recurrent unit (GRU))

Gated Recurrent Unit (GRU), zaman serisi verilerinin işlenmesinde kullanılan bir yinelemeli sinir ağı (RNN) modelidir ve bu çalışmada etkili bir şekilde kullanılmıştır. GRU, özellikle uzun vadeli bağımlılıkları öğrenme yeteneğiyle dikkat çeker. Tez çalışmasının ihtiyaçlarına uygun olarak, GRU'nun daha basit yapısı ve daha az bellek kullanımı, eğitim sürecini hızlandırabilir ve daha düşük hesaplama gücü gereksinimiyle avantaj sağlayabilir. Ayrıca, uzun vadeli bağımlılıkları yönetme ve zaman serisi verilerini işleme konusundaki başarısı, İHA'ların GPS sinyal kaybı durumunda konum tahmininde etkili bir rol oynamasını sağlayabilir.

GRU yapısı, LSTM'e benzer şekilde çalışmakla birlikte daha az sayıda kapı mekanizması içerir. GRU'daki ilk bileşen, güncelleme kapısıdır. Bu kapı, modelin önceki zamandan taşıdığı bilgilerin ne kadarının korunacağını ve yeni gelen veriden ne kadarının dikkate alınacağını belirler. Bu işlem aşağıdaki Denklem (3.15) ile tanımlanır.

$$z_t = \sigma(Wz\, x_t + Uz\, h_{t-1}) \quad (3.15)$$

Burada z_t , güncelleme kapısının çıktısıdır. x_t t zamanındaki giriş verisidir. h_{t-1}, bir önceki zaman adımındaki gizli durumdur. Wz ve Uz ise öğrenilen ağırlık matrisleridir. Ardından gelen reset kapısı, geçmiş bilginin ne kadarının dikkate alınacağını kontrol eder.

Reset kapısı Denklem (3.16) ile hesaplanır. Bu kapı sayesinde, modelin geçmişten gelen bilgiyi sıfırlayıp sıfırlamayacağı belirlenir.

$$r_t = \sigma(Wr\, x_t + Ur\, h_{t-1}) \quad (3.16)$$

Bir sonraki adımda, yeni oluşturulacak aday gizli durumu ($\hat{h}_t$) hesaplanır. Bu hesaplama sırasında, reset kapısının çıktısı kullanılarak önceki gizli durumun ne kadar katkı sağlayacağı kontrol edilir (Denklem (3.17)).

$$\hat{h}_t = tanh(Wh\, x_t + Uh\, (r_t \cdot h_{t-1})) \quad (3.17)$$

Bu ifade, yeni oluşturulacak geçici hafızanın ($\hat{h}_t$) hem güncel girdi hem de sıfırlanmış geçmiş bilgiye bağlı olarak üretildiğini gösterir. Son adımda ise güncel gizli durum h_t, önceki gizli durum h_{t-1} ile yeni aday durum $\hat{h}_t$ arasında bir denge kurularak Denklem (3.18) ile hesaplanır.

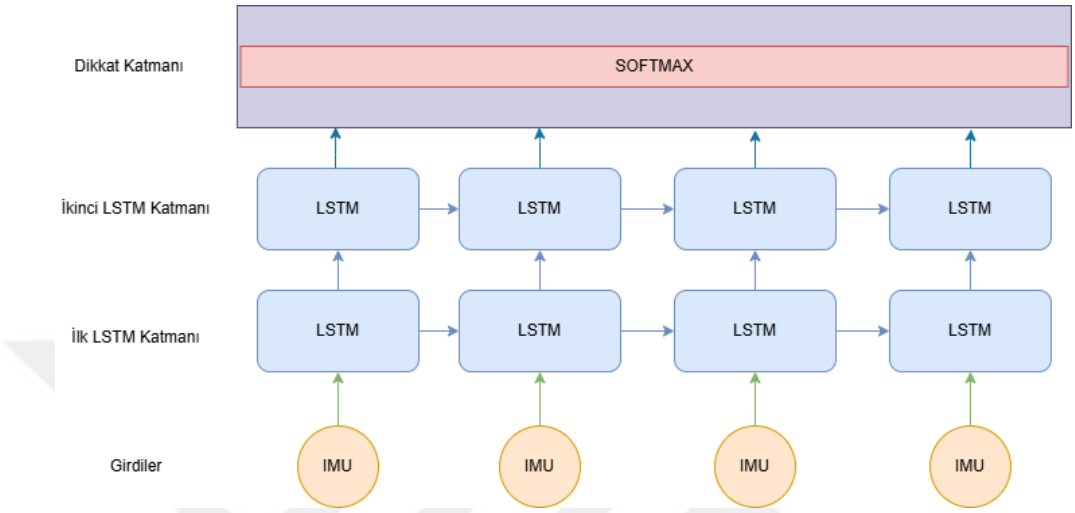
$$h_t = (1 - z_t) \cdot h_{t-1} + z_t \cdot \hat{h}_t \quad (3.18)$$

Bu denklemde, güncelleme kapısının çıktısı olan z_t, geçmiş ve yeni bilgi arasında bir geçiş oranı görevi görür. Sonuç olarak h_t, hem geçmiş bilgiyi hem de güncel girdiyi dengeli bir biçimde yansıtan nihai gizli durumdur.

3.2.7. Dikkat tabanlı hiyerarşik LSTM (Attention-based hierarchical LSTM (AHLSTM))

Attention-based Hierarchical LSTM (AHLSTM) zaman serisi tahminlerinde zamansal ilişkileri daha iyi modellemek için geliştirilmiş bir yapıdır. Model üç ana katmandan oluşur. İlk katman, zaman serisi verisinin zamansal ilişkilerini yakalamak amacıyla bir LSTM ağı kullanır. Bu katman verinin temel dinamiklerini öğrenerek giriş verilerini daha üst seviyede temsil eder. İkinci katman, birinci katmandan gelen çıktılar arasındaki daha karmaşık ilişkileri modellemek için bir başka LSTM ağıdır. Bu katman, sadece zaman serisindeki otokorelasyonları değil aynı zamanda verinin farklı zaman noktaları arasındaki çapraz ilişkileri de öğrenir. Böylece daha derin ve karmaşık yapılar elde edilir. Üçüncü ve son katman ise dikkat mekanizmasıyla çalışır. Dikkat katmanı, ikinci katmandan gelen çıktılar üzerinde çalışarak hangi zaman noktalarının daha önemli

olduğunu öğrenir ve bu noktalara daha fazla ağırlık verir. Bu sayede model özellikle kritik zamansal bilgileri dikkate alarak daha isabetli tahminler yapar. Bu yapı zamansal verilerdeki karmaşık ilişkileri yakalayarak genel tahmin performansını artırır. AHLSTM mimarisine ait yapı şekil 3.4'te gösterilmiştir (Taghizadeh ve Safabakhsh, 2023).



Şekil 3.4. Attention tabanlı hiyerarşik LSTM sistem tasarımı

AHLSTM modelinde ilk olarak giriş verisi x_t, birinci LSTM katmanına aktarılır ve ilgili zaman adımı için gizli durum çıktısı $h_t^{(1)}$ elde edilir. Bu çıktı, üst katmanda daha soyut ilişkilerin öğrenilmesi için temel bir temsil oluşturur (Denklem (3.19)).

$$h_t^{(1)} = LSTM^{(1)}(x_t) \tag{3.19}$$

İkinci LSTM katmanı, ilk katmandan gelen $h_t^{(1)}$ çıktıları üzerine çalışarak zaman serisindeki daha karmaşık yapıları ve üst düzey bağımlılıkları öğrenir. Bu sayede model, temel düzeyde çıkarılan özelliklerin ötesinde, zamanın daha derin bağlamsal ilişkilerini modelleme imkânı bulur. Bu katman, önceki katmandan gelen gizli durumları tekrar işler ve her zaman adımı için yeni bir özellik temsili olan $h_t^{(2)}$ çıktısını üretir (Denklem (3.20)).

$$h_t^{(2)} = LSTM^{(2)}(h_t^{(1)}) \tag{3.20}$$

Son katman olan dikkat (attention) mekanizması, ikinci katmandan gelen $h_t^{(2)}$ çıktıları üzerinde çalışarak hangi zaman adımlarının daha önemli olduğunu belirlemeye

çalışır. Öncelikle her zaman adımı için dikkat skoru e_t hesaplanır. Daha sonra bu skorlar softmax fonksiyonu ile normalize edilerek dikkat ağırlıkları α_t elde edilir. Son adımda, bağlamsal vektör c, bu dikkat ağırlıkları ile $h_t^{(2)}$ çıktılarının ağırlıklı toplamı olarak hesaplanır. Bu vektör, modelin çıktısı olarak daha anlamlı ve odaklı bir temsili sunar (Denklem (3.21)-Denklem (3.23)).

$$e_t = tanh(Wa\, h_t^{(2)} + ba) \quad (3.21)$$

$$\alpha_t = exp(e_t) \,/\, \Sigma\, exp(e_i) \quad (3.22)$$

$$c = \Sigma\, \alpha_t\, h_t^{(2)} \quad (3.23)$$

3.3. Önerilen Yöntemler

3.3.1. Derin dikkat tabanlı görsel ataletsel konumlandırma sistemi

Tez kapsamında yenilikçi bir yöntem olan Derin Dikkat Tabanlı Görsel Ataletsel Konumlandırma Sistemi önerilmiştir. Bu sistemde görsel ataletsel veriler derin dikkat tabanlı modeller ile çıkartılıp bir derin öğrenme modeli ile birleştirilerek konumlandırma yapılmaktadır. Dikkat tabanlı modeller tüm özellikler yerine en ilgili özellikleri vurgulayarak konumlandırma performansını önemli oranda artırmaktadır.

Görsel özellikleri çıkarmak için Bölüm 3.2.1' de tanıtılan dikkat tabanlı CNN mimarisi önerilmiştir. Dikkat tabanlı mimari görüntüdeki en önemli pikselleri yakalayarak iyileştirilmiş görsel özellikleri oluşturmaktadır. Dikkat tabanlı mimari literatürde yaygın olarak kullanılan ResNet50 modeline dikkat mekanizmaları eklenerek oluşturulmuştur.

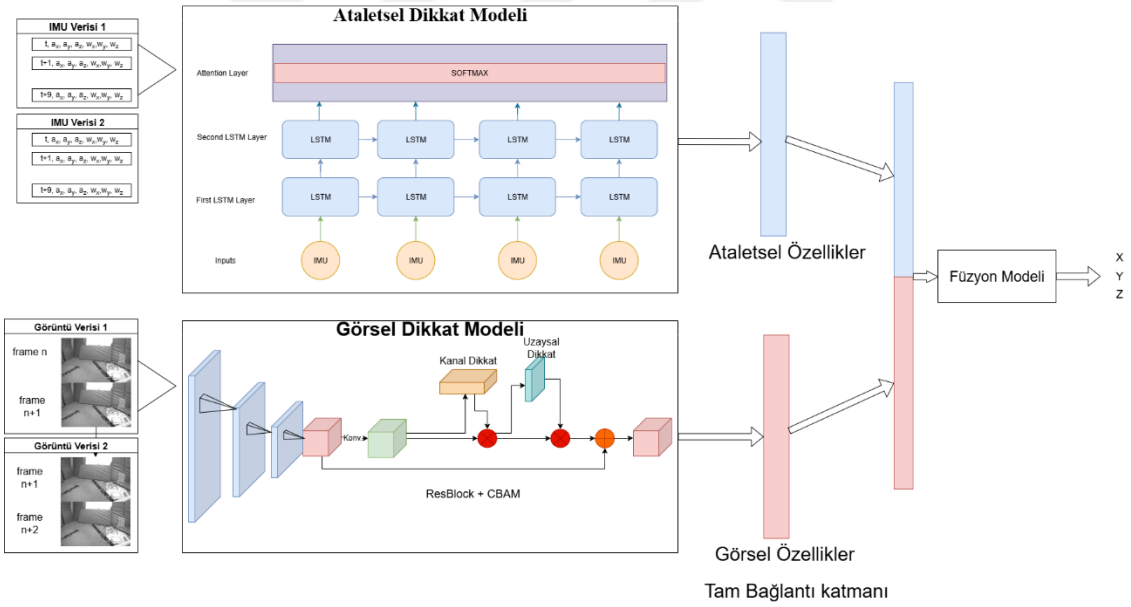
Ataletsel özellikleri çıkarmak için Bölüm 3.3.2'de tanıtılan ataletsel konumlandırma sisteminde en yüksek performansı gösteren ve Bölüm 3.2.7'de detaylı olarak anlatılan Dikkat Tabanlı Hiyerarşik Uzun Kısa Süreli Bellek (Attention-based Hierarchical Long Short-Term Memory (AHLSTM)) modeli kullanılmıştır. Bu model daha önceden GNSS kaybından sensör verileri ile konumlandırma yapmak için kullanılmış ve başarılı olmuştur. Ayrıca tez çalışması kapsamında önerilen ataletsel konumlandırma sisteminde diğer modeller ile performansı karşılaştırılmıştır.

Dikkat tabanlı görsel ataletsel modellerden elde edilen özellikler bir derin öğrenme modeli ile birleştirilerek konumlandırma çıktıları sağlanmıştır. Derin öğrenme

modeli 100 adet görsel 100 adet ataletsel olmak üzere 200 özellik girdisi alarak tahmin işlemini gerçekleştirmektedir.

Önerilen Konumlandırma Sistemini test etmek için Bölüm 3.1.'de tanıtılan EuRoc MAV veri seti kullanılmıştır. Bu veri setinde ataletsel ölçümler olarak jiroskop ve ivmeölçer içeren IMU ile sağlanmıştır. Görsel veriler ise stereo kamera ile sağlanmıştır. IMU verileri 200 HZ, görsel veriler 20 HZ olduğu için model eğitimi ve kullanımında verilerin hizalanması gerekmektedir. Hizalama için iki ardışık görüntü ve bu görüntüler arasında geçen sürede hesaplanan 10 adet IMU verisi modele verilmiştir. Ayrıca genel sistemin performansını iyileştirmek ve CNN hızını artırmak için görüntü verilerinde ¼ oranında olacak şekilde yeniden boyutlandırılmıştır. Hizalama ve yeniden boyutlandırma işlemleri sonrasında her bir eğitim verisi için ataletsel modeli girdi boyutu 10x6, görsel model girdi boyutu 120x188x2 olmuştur.

Önerilen Derin Dikkat Tabanlı Konumlandırma Sistemine ait mimari tasarım Şekil 3.5'te verilmiştir.

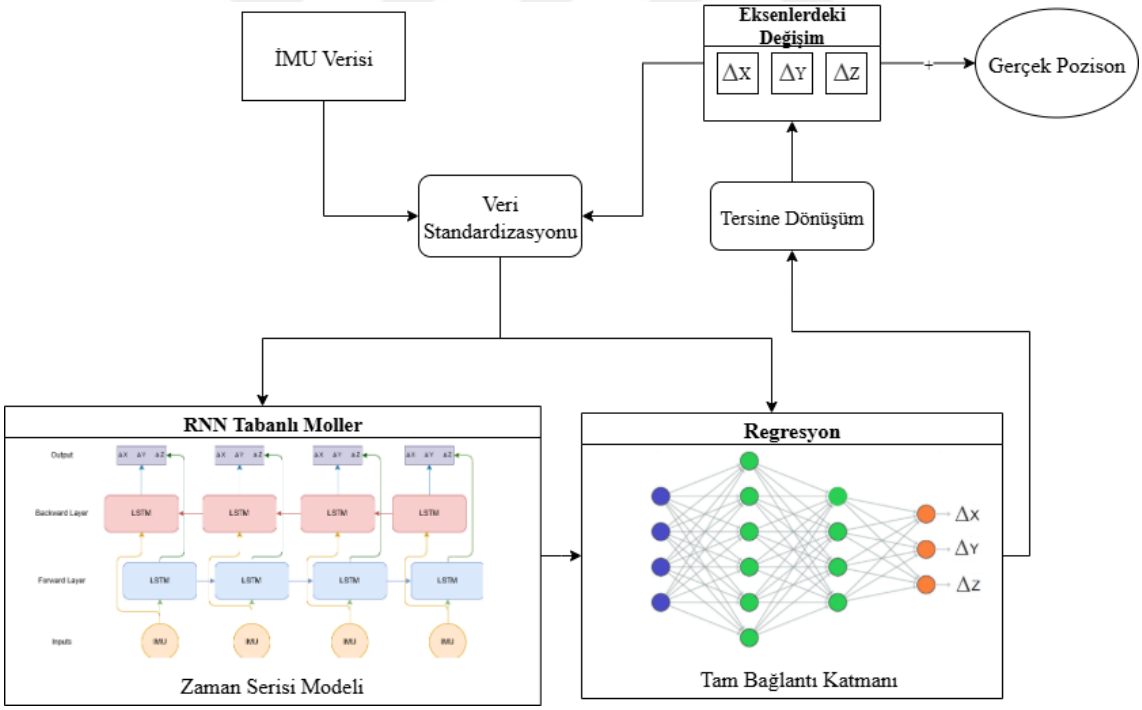


Şekil 3.5. Derin Dikkat Tabanlı Görsel Ataletsel Konumlandırma Sistemi

3.3.2. Derin öğrenme tabanlı ataletsel konumlandırma sistemi

Derin Öğrenme Tabanlı Ataletsel Konumlandırma Sistemi IMU verilerini kullanarak konumlandırma problemlerine yenilikçi bir yaklaşım sunan bir algoritmadır. Geleneksel yöntemlerden farklı olarak, bu model doğrudan konumu tahmin etmeye çalışmak yerine, X, Y ve Z eksenlerindeki değişimleri (ΔX, ΔY, ΔZ) tahmin etmeye

odaklanır. Böylece konumlandırma problemini daha yönetilebilir alt parçalara bölerek daha hassas ve esnek çözümler sunar. Modelin temelinde IMU'nun zaman serisi verilerini işlemek için tasarlanmış RNN bulunur. İlk olarak IMU'dan alınan ivme ve açısal hız verileri standartlaştırma işleminden geçirilir. Böylece tüm veriler aynı ölçekte işlenebilir hale gelir. Standartlaştırılan bu veriler RNN modeline beslenir. RNN zaman içindeki dinamikleri öğrenerek eksenlerdeki hareket değişikliklerini yakalar. RNN'nin ürettiği ara çıktılar tam bağlantılı bir katmana geçirilir ve burada bir regresyon analizi uygulanarak eksenlerdeki değişimler tahmin edilir. Bu yaklaşım doğrudan konum tahmini yapmanın karmaşıklığını azaltır, hareketin detaylı dinamiklerini yakalar ve genel model performansını artırmaktadır. Önerilen yöntem özellikle dinamik ortamlar ve yüksek doğruluk gerektiren uygulamalarda üstün performans göstermektedir. Geleneksel yaklaşımların ötesine geçerek eksenel değişimleri odak noktası haline getirmesi, bu modeli inovatif ve etkili bir çözüm haline getirmektedir. Önerilen yönteme ait mimari Şekil 3.6' da verilmiştir.



Şekil 3.6. Derin Öğrenme Tabanlı Ataletsel Konumlandırma Sistemi

Önerilen yöntemde uygulanan veri standardizasyonu IMU verilerinin çeşitli özelliklerinin (ivme ve açısal hız) aynı ölçeklere getirilmesi işlemidir. Bu süreç makine öğrenmesi modellerinin performansını optimize etmek için kritik bir adımı teşkil etmektedir. IMU verilerindeki farklı özellikler genellikle birbirinden farklı birim ve

ölçeklere sahip olabilir. Bu veriler standardize edilmeden doğrudan modele aktarılırsa, model bazı özelliklere diğerlerine kıyasla daha fazla ağırlık verebilir ve bu da sonuçların doğruluğunu olumsuz yönde etkileyebilir.

Standardizasyon her bir özelliğin ortalamasını sıfıra standart sapmasını bire getirerek verilerin ölçeklenmesini sağlar. Böylece tüm özellikler aynı düzeyde önem kazanır ve modelin öğrenme süreci dengelenir. Veri standardizasyonu özellikle zaman serisi verileri ve IMU verileri gibi dinamik ve çok yönlü özellikler içeren veri kümelerinde yaygın olarak kullanılır. Bu formül her bir veri noktasını kendi özelliğinin ortalama ve standart sapmasına göre yeniden ölçeklendirir. Sonuç olarak, her bir özellik için veriler sıfır ortalama ve birim standart sapmaya sahip olur. Standardizasyon x ham veri değerini, μ ortalama değeri ve σ standart sapmayı temsil edecek şekilde Denklem (3.24)'te ifade edilmiştir.

$$x_{standardize} = \frac{x - \mu}{\sigma} \tag{3.24}$$

4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA

4.1. Deneysel Çalışma Parametreleri ve Değerlendirme Metrikleri

EuRoc MAV veri seti ile Derin Dikkat Tabanlı Görsel Ataletsel Konumlandırma Sistemi eğitilmiş ve literatür karşılaştırması yapılmıştır. Eğitim için tüm veriden karıştırma (shuffle) işlemi ile beraber %80 eğitim %20 test verisi seçilmiştir. Hem literatür karşılaştırmasında hem de eğitimde Hataların Karesinin Ortalamasının Karekökü (Root Mean Square Error (RMSE)) metriği kullanılmıştır. Optimizasyon algoritması olarak Adaptif Momentum (ADAM) algoritması kullanılmıştır. Tüm oda ve zorluk seviyeleri için 250'şer adım eğitim yapılmıştır. Eğitim işlemi NVIDIA RTX A2000 (12GB) ekran kartı üzerinde gerçekleştirilmiştir.

Webots simülasyon verisi ile konumlandırma eğitimi için ilk olarak veriler birleştirilmiş ve normalize edilmiştir. Daha sonra %80 eğitim %20 test verisi olarak ayrılmıştır. Hata fonksiyonu olarak Ortalama kare hatası (MSE) kullanılmıştır. Optimizasyon algoritması olarak Adaptif Momentum (ADAM) algoritması kullanılmıştır. Eğitim aşamasında dropout ve dropout olmadan iki adet 250'şer adım eğitim yapılmıştır. Eğitimlerdeki dropout değeri 0,2 olarak kullanılmıştır. Eğitim için donamım olarak Nvidia RTX3060 (6GB) ekran kartı kullanılmıştır.

Üçüncü eğitimde Özellik çıkarıcı sistemin EuRoc MAV veri setinde tam konumlandırma yeteneğinin düşük olma sebebi araştırılmıştır. Bunun için simülasyon veri setinin tüm verileri içeren veride, sadece EuRoc MAV içerisinde bulunan özelik verileri ile EuRoc MAV özelliklerine verilerine ek olarak pusulanın bulunduğu veride olmak üzere üç farklı eğitim yapılmıştır. Eğitim işlemi için veri seti %80 eğitim %20 test verisi olarak ayrılmıştır. ADAM optimizasyon algoritması, 0.2 dropout değeri ile 250'şer adım eğitilmiştir.

Modelin performansını değerlendirmek için çeşitli regresyon metrikleri kullanılmıştır. Bunlardan ilki, Denklem (4.1) ile hesaplanan tahmin edilen değerler ile gerçek değerler arasındaki farkların karesinin ortalamasını alan Ortalama Kare Hata (Mean Squared Error – MSE)'dır. Bu metrik, büyük hataları daha fazla cezalandırarak modelin genel doğruluğunu değerlendirir. İkinci olarak Denklem (4.2) ile hesaplanan Ortalama Mutlak Hata (Mean Absolute Error – MAE), tahmin hatalarının mutlak değerlerinin ortalamasını alır ve MSE'ye göre daha az duyarlıdır. MSE'nin karekökü alınarak elde edilen (Denklem 4.3) ile hesaplanan Kök Ortalama Kare Hata (Root Mean

Squared Error – RMSE) ise yorumlanabilirliği artırmak için kullanılır ve birim olarak hedef değişkenle aynı düzeye sahiptir. Modelin açıklayıcılık gücünü ölçmek için ise Denklem (4.4) ile ifade edilen Belirleme Katsayısı (R^2) kullanılmıştır; bu metrik, modelin toplam varyansı ne ölçüde açıkladığını gösterir. Son olarak, tahmin hatalarının yüzde cinsinden ne kadar saptığını ölçen Ortalama Mutlak Yüzde Hata (Mean Absolute Percentage Error – MAPE) metriği hesaplanmıştır (Denklem (4.5)). Tüm bu metrikler, modelin hem genel doğruluğunu hem de uygulamaya uygunluğunu değerlendirmede önemli rol oynamaktadır.

$$MSE = \sum_i^n \frac{(y_i - \hat{y}_i)^2}{n} \quad (4.1)$$

$$MAE = \frac{1}{n} \sum_i^n |y_i - \hat{y}_i| \quad (4.2)$$

$$RMSE = \sqrt{\sum_i^n \frac{(y_i - \hat{y}_i)^2}{n}} \quad (4.3)$$

$$R^2 = 1 - \frac{\Sigma(y_i - \hat{y}_i)^2}{\Sigma(y_i - \bar{y})^2} \quad (4.4)$$

$$MAPE = \frac{1}{n} \sum_i^n \left(\left| \frac{y_i - \hat{y}_i}{y_i} \right| \right) * 100\% \quad (4.5)$$

4.2. Derin Öğrenme Tabanlı Ataletsel Konumlandırma Sistemi Testi

Yapılan çalışmalara ait dropout ve dropout içermeyen eğitimler için elde edilmiş sonuçlar Tablo 4.1'de verilmiştir. Metriklere göre en iyi performansı gösteren model koyu renkte işaretlenmiştir.

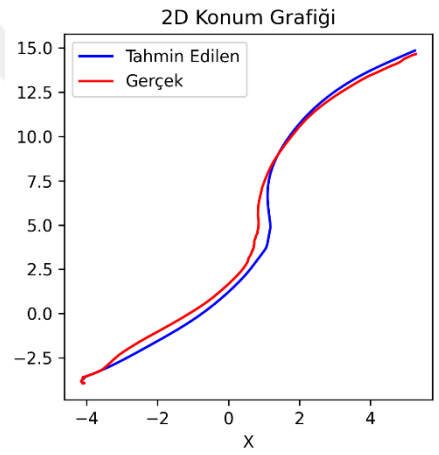
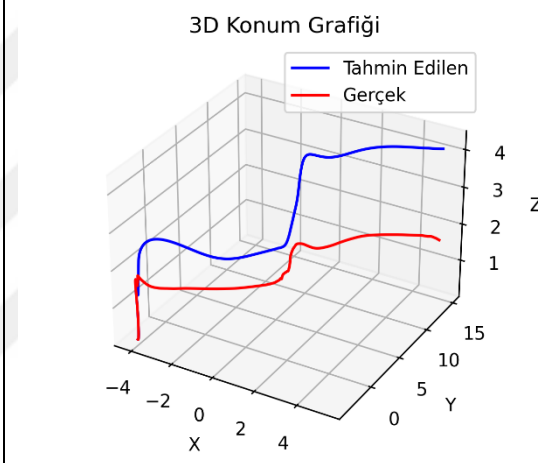
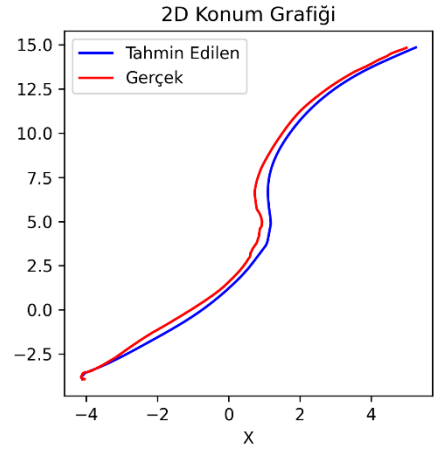
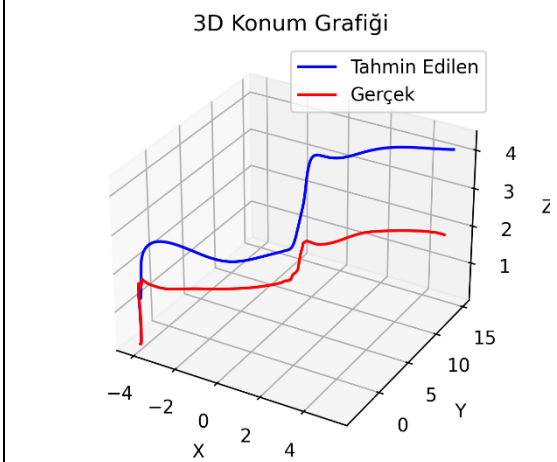
Tablo 4.1. Eğitilen modelin hata çıktıları

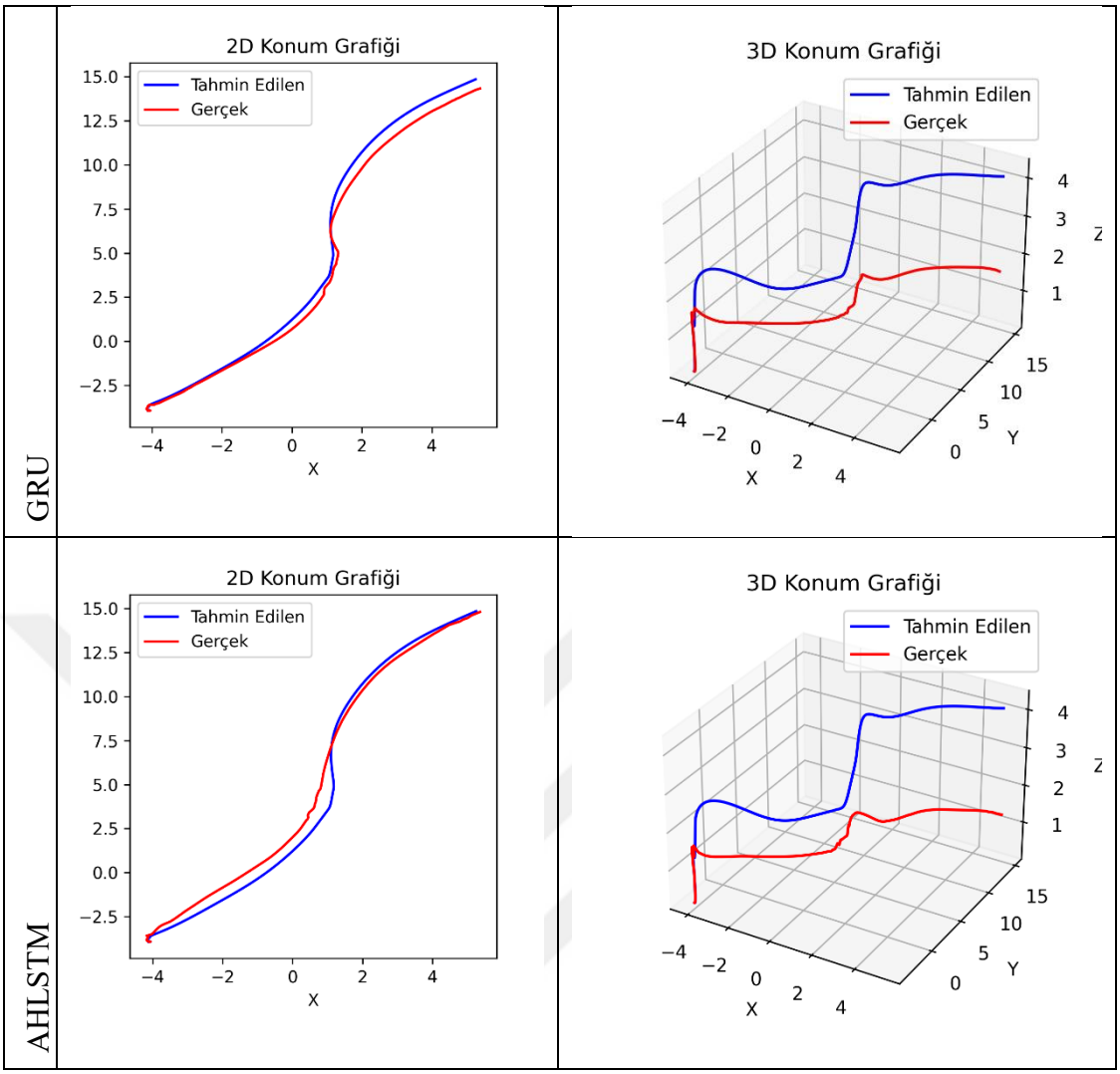
	Dropout İçeren Model					**Dropout İçermeyen Model**				
	MSE	**MAE**	**RMSE**	**R²**	**MAPE**	**MSE**	**MAE**	**RMSE**	**R²**	**MAPE**
DNN	0.1649	0.2236	0.4061	0.8346	1.1295	0.0967	0.1197	0.3110	0.9027	0.5300
RNN	0.0068	0.0487	0.0826	0.9931	0.2536	0.0043	0.0339	0.0656	0.9956	0.1679
LSTM	0.0011	0.0147	0.0329	0.9989	0.0719	0.0012	0.0114	0.0349	0.9987	0.0464
BİLSTM	0.0022	0.0169	0.0473	0.9977	0.0953	0.0012	0.0111	0.0349	0.9987	0.0519
GRU	0.0016	0.0183	0.0401	0.9983	0.0958	0.0011	0.0133	0.0344	0.9988	0.0539
AHLSTM	**0.0008**	**0.0104**	**0.0287**	**0.9991**	**0.0525**	**0.0010**	**0.0080**	**0.0318**	**0.9989**	**0.0366**

Tablo 4.1'e göre konumlandırma için önerilen modellerin önemli derecede başarılı olduğu görülmektedir. Eğitime eklenen dropout genel olarak modellerde aşırı öğrenmeyi azaltmakta etkili olmuştur, ancak bazı modellerde hataları artırmıştır. Eğitim çıktılarına göre en iyi performansı gösteren model dropout içeren AHLSTM modelidir. AHLSTM tüm metriklere göre en iyi performansı göstermiştir.

Yapılan eğitim sonucunda en iyi performansı gösteren modellere ile test verisi üzerinde yapılan deneylere ait çıktılar Tablo 4.2'de verilmiştir.

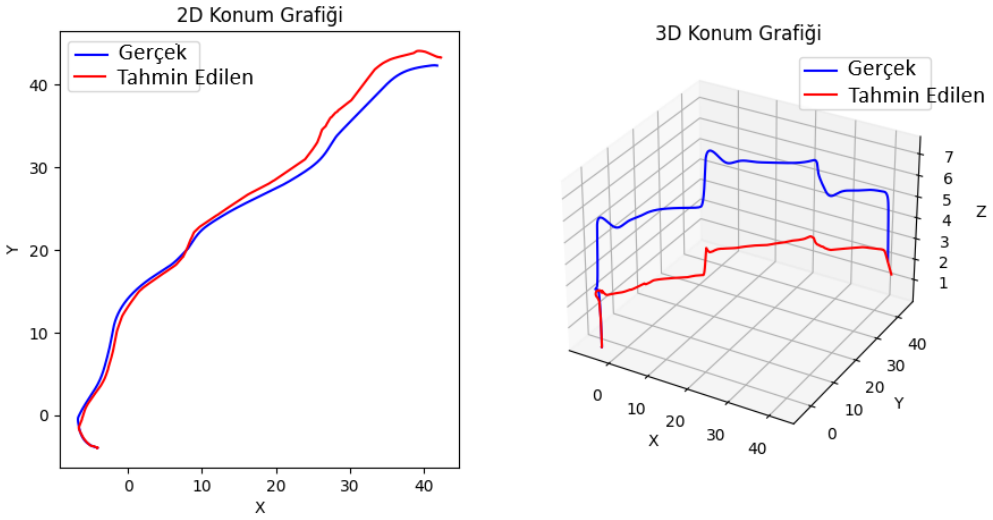
Tablo 4.2. Önerilen model konumlandırma çıktıları

	Test Verisi	
	2D	3D
LSTM		
BİLSTM		



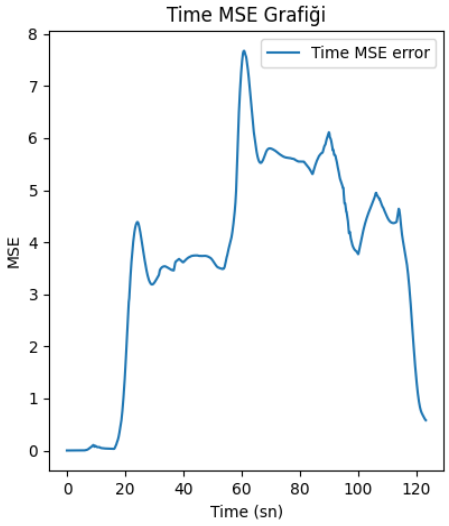
Tablo 4.2'de verilen çıktılar incelendiğinde önerilen yöntem ile farklı modeller kullanılarak başarılı konumlandırma yapılabileceği görülmüştür.

Önerilen sistemin konumlandırma için zaman içerisinde oluşan hatası incelenmiştir. Bu inceleme için eğitim verisinde bulunmayan bir uçuş dropout içeren AHLSTM modeli ile test edilmiştir. Uçuşa ait gerçek ve tahmin edilen çıktılar şekil 4.1' de verilmiştir.



Şekil 4.1. Önerilen sistemin konumlandırma çıktıları

Yapılan uçuşun zamana bağlı RMSE metriğine göre gerçek veriler ve tahmin edilen veriler arasındaki hata şekil 4.2'de verilmiştir.



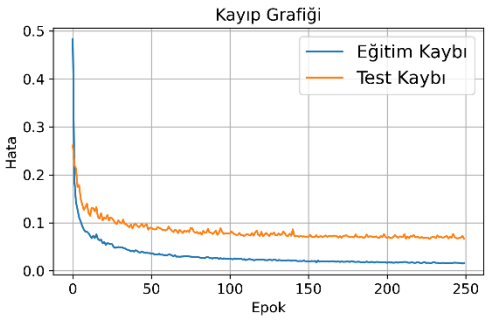
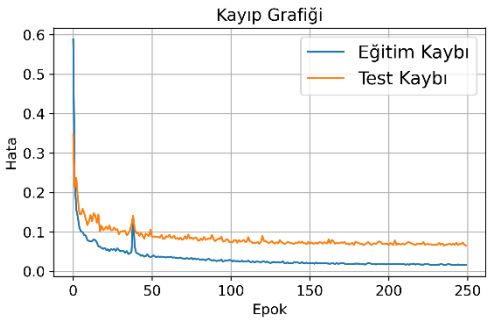
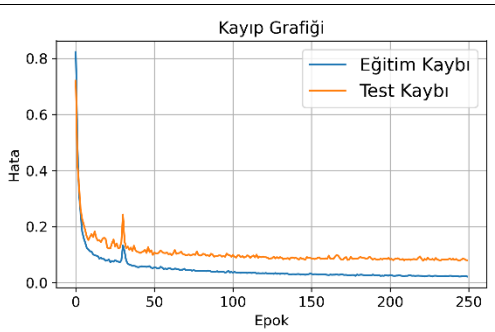
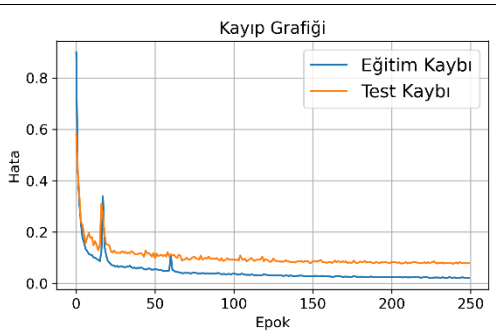
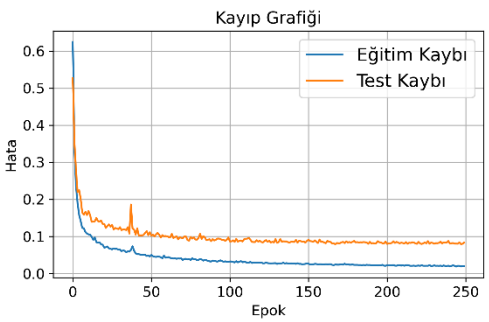
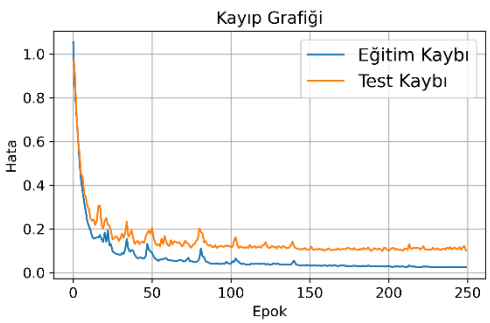
Şekil 4.2. Önerilen sistemin zamana bağlı konumlandırma hatası

Gerçekleştirilen uçuş ve hata grafiği incelendiğinde özellik çıkarıcı sistemden farklı olarak hata grafiğinin sürekli artmak yerine zamanla dalgalandığı gözlemlenmiştir. Şekil 4.1'de verilen uçuş yörüngesi ve şekil 4.2'de verilen zaman bağlı hata incelendiğinde bu hata oranının İHA'nın yükselme alçalma gibi farklı durumlara göre girdilerden etkilendiği tespit edilmiştir. Sonuç olarak tam konumlandırıcı sistemin başarısının eksenlerdeki hareketlere göre değişebileceği gözlemlenmiştir.

4.3. Derin Dikkat Tabanlı Görsel Ataletsel Konumlandırma Sistemi Testi

Bu bölümde önerilen Derin Dikkat Tabanlı Görsel Ataletsel Konumlandırma Sisteminin eğitim sonucunda elde edilen çıktıları incelenmiştir. Elde edilen çıktılar literatürdeki diğer çalışmalar ile karşılaştırılmıştır. Çalışma kapsamında elde edilen eğitim ve test kaybı eğrileri, farklı zorluk seviyeleri (Easy, Medium, Difficult) ve iki ayrı deney ortamı (Vicon Room 1 ve Vicon Room 2) dikkate alınarak değerlendirilmiştir. Tüm oda ve zorluk koşullarında yapılan eğitime ait kayıp grafikleri Tablo 4.3'te sunulmuştur.

Tablo 4.3. Eğitim Hata Grafikleri

	Vicon Room 1	Vicon Room 2
Easy		
Medium		
Difficult		

Kolay (Easy) zorluk seviyesinde her iki ortamda da eğitim ve test kayıplarının hızlı bir şekilde azaldığı ve düşük seviyelerde sabitlendiği gözlemlenmiştir. Bu durum modelin basit görevlerde yüksek doğrulukla öğrenebildiğini ve genelleme kabiliyetinin yeterli olduğunu göstermektedir.

Orta (Medium) seviyede ise kayıplarda zaman zaman dalgalanmalar gözlemlense de eğitim sürecinin ilerleyen adımlarında daha stabil bir seyir izlediği görülmektedir. Bu, modelin daha karmaşık örneklerde öğrenme sürecine uyum sağladığını ancak bu süreçte daha fazla varyansla karşılaştığını ortaya koymaktadır.

Zor (Difficult) seviyede ise özellikle Vicon Room 2 ortamında test kayıplarının daha yüksek seviyelerde seyrettiği ve dalgalanmaların belirginleştiği dikkat çekmektedir. Bu durum artan görev karmaşıklığının modelin öğrenme sürecini zorlaştırdığını ve genelleme kapasitesini sınırladığını göstermektedir.

Ortamlar arası karşılaştırma yapıldığında ise Vicon Room 1'deki sonuçların, özellikle orta ve zor seviyelerde, daha istikrarlı olduğu ve modelin daha kararlı bir şekilde öğrenme gerçekleştirdiği anlaşılmaktadır. Buna karşın, Vicon Room 2 ortamında gözlemlenen yüksek varyans ve test kaybındaki dalgalanmalar, bu ortamın daha fazla gürültü içerdiğini veya deneysel koşulların modele daha az uygun olduğunu düşündürmektedir.

Sonuç olarak deneysel bulgular modelin performansının yalnızca görev zorluk derecesine değil aynı zamanda deney ortamının özelliklerine de duyarlı olduğunu ortaya koymaktadır. Özellikle karmaşık görevlerde ve değişken ortam koşullarında modelin genelleme yeteneğini artırmaya yönelik düzenleme (regularization) yöntemlerinin ve veri çeşitliliğini artıracak stratejilerin kullanılması faydalı olabilme potansiyeline sahiptir.

Önerilen yöntemin performansı, literatürdeki görsel-atalet odometri yöntemleri ile kapsamlı bir şekilde karşılaştırılmıştır. Değerlendirme altı farklı test veri seti (V101, V102, V103, V201, V202, V203) üzerinde gerçekleştirilmiş ve performans metriği olarak RMSE kullanılmıştır. Önerilen yöntemin diğer çalışmalar ile karşılaştırması Tablo 4.4'te verilmiştir. Koyu renkle işaretlenen değerler en düşük hata değerini, "-" sembolü ise ilgili veri üzerinde makalede sonuç bulunmadığını belirtmektedir.

Tablo 4.4. RMSE metriği ile literatür karşılaştırması

		V101	V102	V103	V201	V202	V203	avg
Klasik VIO	VINS-Mono_loop (2018)	0.068	0.084	0.190	0.081	**0.016**	0.220	0.110
	VINS-Mono (2018)	0.079	0.110	0.180	0.080	0.160	0.270	0.147
	ORB-SLAM3 (2020)	0.049	**0.015**	0.037	0.042	0.021	**0.027**	**0.032**
	Kimera-VIO (2020)	0.050	0.080	0.070	0.080	0.100	0.210	0.098
	PL-VINS (2022)	-	0.092	0.152	-	-	0.182	0.142
	OKVIS2 (VIO) (2022)	0.043	0.037	0.036	0.044	0.044	0.063	0.045
	HybVIO(mono) (2022)	0.039	0.044	0.065	0.044	0.047	0.056	0.049
Öğrenme Tabanlı VIO	Baldini ve diğ., (2020)	0.030	0.050	0.100	0.050	0.050	0.080	0.060
	VIUNet (2023)	-	0.080	0.090	0.070	0.080	0.090	0.082
	Adaptive VIO (2024)	0.035	0.045	0.073	0.052	0.086	-	0.058
	Mix-VIO (2024)	0.063	0.097	0.150	0.063	0.070	0.120	0.094
	VIIONet (2022)	0.036	0.070	0.075	0.0302	0.063	0.125	0.066
	HVIOnet (2022)	0.110	0.087	0.190	0.0530	0.183	0.183	0.134
	VIO-DualProNet (2024)	0.050	0.080	0.110	0.0500	0.070	0.180	0.090
	Önerilen Yöntem	**0.018**	0.030	**0.034**	**0.016**	0.030	0.067	**0.032**

EuRoC veri kümesi üzerinde gerçekleştirilen değerlendirmede, altı farklı sekans (V101, V102, V103, V201, V202, V203) kullanılarak RMSE metrikleri ölçümlenmiştir. Önerilen yöntem 0.032 ortalama hata ile genel performansta en iyi sonuçlardan birini elde etmiştir ve literatürdeki yöntemlerle rekabet edebilir düzeyde performans sergilemiştir.

Klasik VIO yöntemleri arasında VINS-Mono (2018) ve ORB-SLAM3 (2020) gibi yaygın kullanılan algoritmalar incelendiğinde, farklı performans seviyeleri gözlenmektedir. VINS-Mono'nun iki farklı konfigürasyonu sırasıyla 0.110 ve 0.147 ortalama hata değerleri sergilemiştir. Bu değerler önerilen yönteme kıyasla belirgin şekilde yüksek hata oranlarını işaret etmektedir. ORB-SLAM3 ise 0.032 ortalama hata ile klasik yöntemler arasında en iyi performansı göstermiş ve önerilen yöntemle çok yakın sonuçlar elde etmiştir. Kimera-VIO (2020) 0.098 ortalama hata ile orta düzeyde bir performans sergilemiş, PL-VIO (2022) ise 0.142 hata değeri ile bu kategoride en düşük performansı göstermiştir.

Derin öğrenme tabanlı VIO yöntemleri genel olarak klasik yöntemlere kıyasla değişken performans sergilemiştir. OKVIS2 (VIO) (2022) 0.045 ve HybVIO (mono) (2022) 0.049 ortalama hata değerleri ile makul sonuçlar elde etmiş ancak önerilen yöntemin performansına ulaşamamıştır. Baldini ve diğerleri tarafından önerilen yöntem 0.060 ortalama hata ile bu kategoride orta seviyede performans göstermiştir. VIONet (2023) 0.082 hata değeri ile derin öğrenme yaklaşımları arasında görece yüksek hata oranı sergilemiştir. Bu durum derin öğrenme tabanlı yöntemlerin her zaman üstün performans sağlamadığını göstermektedir.

Son yıllarda önerilen hibrit ve adaptif VIO yöntemlerinin performans analizi bu alandaki çeşitliliği ve gelişim potansiyelini ortaya koymaktadır. Adaptive VIO (2024) 0.058 ortalama hata ile güncel yöntemler arasında en iyi performansı sergilemiş, Mix-VIO (2024) 0.094 hata değeri ile orta düzeyde sonuçlar elde etmiştir. VIIONet (2022) 0.066 hata oranı ile hafif ağ yapıları açısından makul performans göstermiş, HVIONet (2022) ise 0.134 hata ile bu kategoride en yüksek hata oranını kaydetmiştir. VIO-DualProNet (2024) 0.090 hata değeri ile dual network yaklaşımının sınırlarını işaret etmektedir.

Önerilen yöntem 0.032 ortalama hata değeri ile literatürdeki en iyi sonuçları elde eden yöntemler arasında yer almaktadır. Sekans bazlı performans analizi incelendiğinde, özellikle V201 sekansında 0.016 hata değeri ile tüm yöntemler arasında en düşük hatayı kaydetmiştir. V102 sekansında 0.030 hata değeri ile ORB-SLAM3'ün 0.015 değerine yakın performans sergilemiş, V103 (0.034) ve V203 (0.067) sekanslarında ise tutarlı sonuçlar elde etmiştir. Bu veriler, önerilen yöntemin farklı hareket dinamikleri ve çevresel koşullarda stabil performans sergilediğini göstermektedir.

Sekans bazlı detaylı performans incelemesi, önerilen yöntemin gürbüzlük (robust) karakteristiklerini ortaya koymaktadır. V101 sekansında önerilen yöntem 0.018 hata değeri ile ORB-SLAM3'ün 0.049 değerinden belirgin şekilde üstün performans sergilemiştir. V202 sekansında 0.030 hata değeri, genel literatür ortalamasının altında kalarak tutarlı sonuçları desteklemektedir. Bu bulgular, algoritmanın değişken çevresel koşullar ve farklı hareket paternleri karşısında güvenilir performans sergilediğini işaret etmektedir.

Deneysel sonuçların istatistiksel değerlendirmesi, önerilen VIO yönteminin literatürdeki en iyi performans gösteren algoritmalarla rekabet edebilir düzeyde sonuçlar elde ettiğini göstermektedir. Özellikle klasik yöntemlerden ORB-SLAM3 ile kıyaslandığında minimal performans farkı bulunmakta, bu fark sadece 0.001 değerindedir. Derin öğrenme tabanlı yöntemlerle kıyaslandığında ise önerilen yaklaşım %31-78 arasında performans iyileştirmesi sağlamıştır. Bu sonuçlar, metodolojik yaklaşımın etkinliğini kanıtlamakta ve VIO alanında anlamlı bir ilerlemeyi temsil etmektedir.

Sonuç olarak karşılaştırmalı performans analizi, önerilen yöntemin VIO literatüründe önemli bir konuma sahip olduğunu ortaya koymaktadır. Elde edilen 0.033 ortalama hata değeri mevcut state-of-the-art yöntemlerle rekabet edebilir seviyede olup, özellikle gürbüzlük ve tutarlılık açısından literatüre anlamlı bir katkı sağlamaktadır.

Farklı sekanslar üzerindeki tutarlı performans, algoritmanın pratik uygulamalarda güvenilir bir alternatif oluşturduğunu göstermekte ve gelecekteki VIO araştırmaları için gürbüz bir temel sunmaktadır.

Önerilen yöntemin performansı literatürdeki VIO yöntemleri ile hata oranlarının sıralamasına göre karşılaştırılmıştır. Yapılan karşılaştırmaya Tablo 4.5'te verilmiştir. Koyu renkle işaretlenen değerler en düşük hata değerini, "-" sembolü ise ilgili veri üzerinde makalede sonuç bulunmadığını belirtmektedir.

Tablo 4.5. Sıralama bazlı karşılaştırma

		V101	V102	V103	V201	V202	V203	avg
Klasik VIO	VINS-Mono_loop (2018)	11	11	14	14	1	13	9.250
	VINS-Mono (2018)	12	15	13	12	13	14	13.167
	ORB-SLAM3 (2020)	7	1	3	3	2	1	2.833
	Kimera-VIO (2020)	8	8	5	12	12	12	9.500
	PL-VINS (2022)	-	13	12	-	-	10	11.667
	OKVIS2 (VIO) (2022)	6	3	2	4	4	3	2.667
	HybVIO(mono) (2022)	5	4	4	4	5	2	4.000
Öğrenme Tabanlı VIO	Baldini ve diğ., (2020)	2	6	9	6	6	5	5.667
	VIUNet (2023)	-	8	8	11	10	6	8.600
	Adaptive VIO (2024)	3	5	6	8	11	-	6.600
	Mix-VIO (2024)	10	14	11	10	8	7	10.000
	VIIONet (2022)	4	7	7	2	7	8	5.833
	HVIOnet (2022)	13	12	14	9	14	11	12.167
	VIO-DualProNet (2024)	8	8	10	6	8	9	8.167
	Önerilen Yöntem	1	2	1	1	3	4	**2.000**

Sıralama tablosunda önerilen yöntem, 2.000 ortalama sıralama skoru ile karşılaştırılan tüm yöntemlerden üstün performans sergilemiştir. Bu sonuç, yöntemin farklı veri setlerinde tutarlı bir şekilde en iyi performansları gösterdiğini kanıtlamaktadır. V101, V103 ve V201 veri setlerinde 1. sırada yer alarak mükemmel performans sergilerken, V102'de 2. sıra, V202'de 3. sıra ve V203'te 4. sıra ile tüm senaryolarda ilk beş içerisinde kalmıştır.

En yakın rakip olan OKVIS2 (VIO) yöntemi 2.667 ortalama sıralama ile ikinci sırada yer almaktadır. OKVIS2'nin V103 ve V203'te 2. sıra alması dikkat çekici olsa da V101'de 6. sıra gibi daha zayıf performanslar da sergilemiştir. ORB-SLAM3, 2.833 ortalama ile üçüncü sırada yer alırken, özellikle V102 ve V203'te 1. sıralama elde etmesi güçlü yönlerini ortaya koymaktadır. Bu üç yöntemin ilk üç sırada yer alması, görsel-atalet odometri alanında bu yaklaşımların etkinliğini göstermektedir.

Klasik VIO yöntemleri arasında önemli performans farklılıkları gözlenmektedir. HybVIO (4.000 ortalama) ve Baldini ve diğerleri tarafından önerilen yöntem (5.667

ortalama) makul sıralamalar elde ederken, VINS-Mono serisinin her iki versiyonu da 9.250 ve 13.167 ortalama değerleri ile zayıf performans göstermiştir. Kimera-VIO (9.500) ve PL-VINS (11.667) de benzer şekilde düşük sıralamalar elde etmiştir. Bu durum, klasik yöntemlerin tasarım yaklaşımları ve optimizasyon stratejilerindeki farklılıkların performansa önemli etkisini göstermektedir.

Öğrenme tabanlı yöntemler arasında da geniş bir performans yelpazesi görülmektedir. VIIONet (5.833) ve Adaptive VIO (6.600) görece iyi sıralamalar elde ederken, VIO-DualProNet (8.167) ve VIUNet (8.600) orta seviye performans sergilemiştir. Mix-VIO (10.000) ve özellikle HVIOnet (12.167) bu kategori içerisinde en zayıf sonuçları göstermiştir. Bu kategori içerisindeki geniş performans yelpazesi, öğrenme tabanlı yaklaşımların eğitim stratejileri ve model mimarilerindeki farklılıkların etkisini yansıtmaktadır.

V101 ve V201 veri setlerinde önerilen yöntem birinci sıralarda yer alarak en kolay senaryolarda mutlak üstünlük sağlamıştır. Bu veri setlerinde OKVIS2, VIIONet ve HybVIO gibi yöntemler de üst sıralarda yer alırken, VINS-Mono serisi ve HVIOnet alt sıralarda kalmıştır. V103 gibi orta-zorlu senaryolarda önerilen yöntem yine birinci sırada yer alırken, OKVIS2 ve ORB-SLAM3 yakın performans sergilemiştir. Bu veri setinde öğrenme tabanlı yöntemlerin performansı daha değişken olmuş, bazıları üst sıralarda yer alırken diğerleri alt sıralarda kalmıştır.

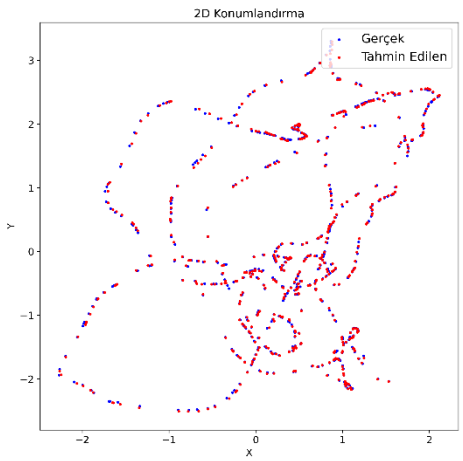
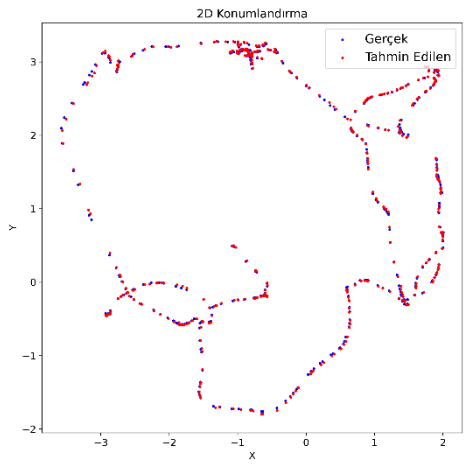
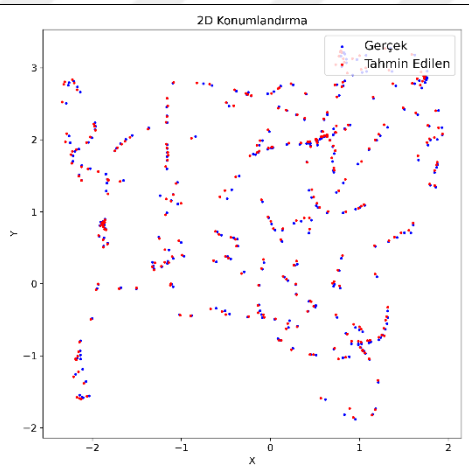
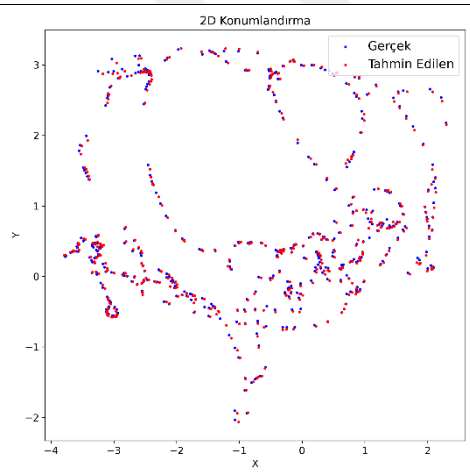
V203 gibi en zorlu senaryolarda ORB-SLAM3'ün birinci sıra alması dikkat çekicidir, ancak önerilen yöntem 4. sıra ile hala üst sıralarda yer almaktadır. Bu zorlu senaryoda genel olarak tüm yöntemlerin performansı daha yakın değerler göstermiş, ancak önerilen yöntem yine de rekabetçi konumunu korumuştur. V202 ve V203 gibi daha zorlu veri setlerinde bile önerilen yöntemin ilk beş içerisinde kalması, yöntemin robustluğunu ve farklı koşullara adaptasyon kabiliyetini ortaya koymaktadır.

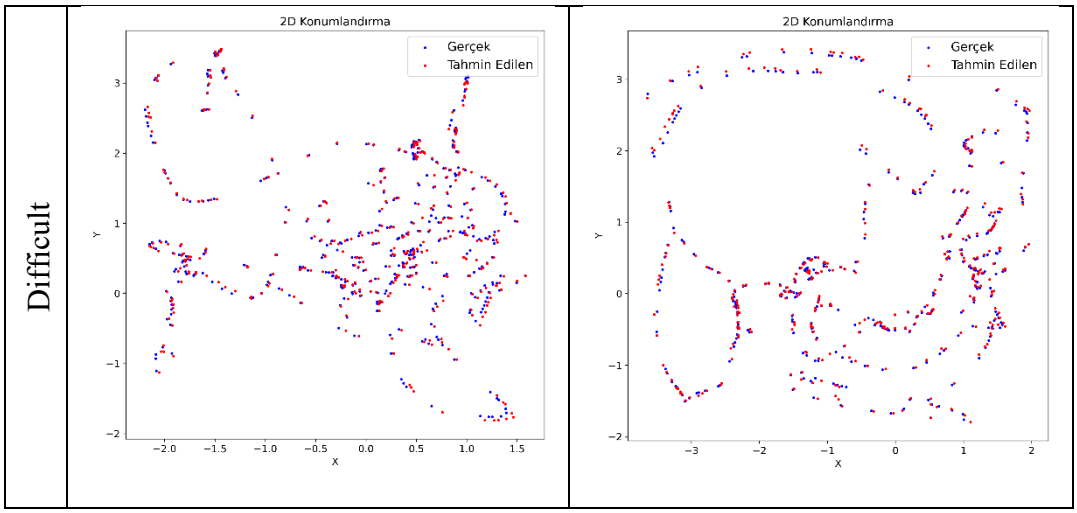
Sıralama tabanlı değerlendirme, önerilen yöntemin görsel-atalet odometri alanında hem tutarlılık hem de doğruluk açısından üstün performans sergilediğini kanıtlamaktadır. 2.000 ortalama sıralama skoru ile elde edilen birinci sıra, yöntemin farklı zorluk seviyelerindeki veri setlerinde güvenilir sonuçlar ürettiğini göstermektedir. Bu başarı, önerilen yöntemin literatüre önemli bir katkı sağladığını ve pratik uygulamalarda güvenilir bir seçenek olduğunu ortaya koymaktadır.

Tablo 4.6‘da 2D konumlandırma grafikleri önerilen görsel-atalet odometri (VIO) yönteminin farklı zorluk seviyelerindeki Vicon Room 1 ve Vicon Room 2 veri setleri üzerindeki performansını detaylı bir şekilde ortaya koymaktadır. Her bir grafikte "Actual"

(Gerçek) etiketli mavi noktalar, zemindeki gerçek konum referansını temsil ederken, "Predicted" (Tahmin Edilen) etiketli kırmızı noktalar, önerilen yöntem tarafından tahmin edilen yörüngeyi göstermektedir. Bu görselleştirmeler, yöntemin doğruluğu ve farklı çevresel koşullara karşı gürbüzlüğü hakkında önemli bilgiler sunmaktadır.

Tablo 4.6. 2D Konumlandırma Grafikleri

	Vicon Room 1	Vicon Room 2
Easy		
Medium		



Kolay zorluk seviyesindeki Vicon Room 1 ve Vicon Room 2 veri setlerinde, önerilen yöntemin tahmin ettiği yörüngeler gerçek yörüngeye son derece yakın bir uyum sergilemektedir. Kırmızı ve mavi noktaların büyük ölçüde üst üste bindiği gözlemlenmektedir; bu durum, yöntemin ideal veya nispeten daha az zorlayıcı koşullar altında dahi yüksek hassasiyetle konum tahmini yapabildiğini doğrulamaktadır. Özellikle yörüngenin genel şekli ve dönüş noktaları gerçek verilerle başarılı bir şekilde eşleşmekte, bu da yöntemin kolay senaryolarda dahi çevresel haritalandırma ve robotik navigasyon gibi uygulamalar için yeterli doğrulukta konum bilgisi sağlayabildiğini göstermektedir.

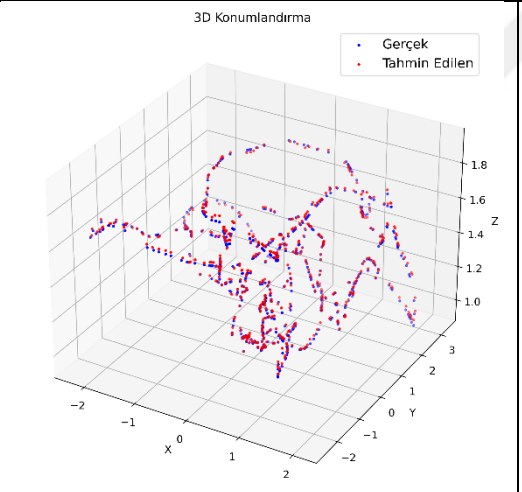
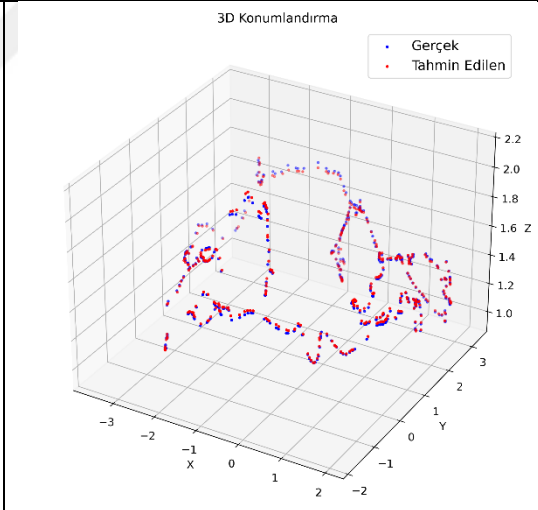
Orta zorluk seviyesine geçildiğinde, Vicon Room 1 ve Vicon Room 2 veri setlerinde de önerilen yöntemin performansı tutarlılığını korumaktadır. Tahmin edilen yörüngeler hala gerçek yörüngeyi yakından takip etmekte ancak bazı bölgelerde kırmızı ve mavi noktalar arasında hafif sapmaların başladığı fark edilmektedir. Bu sapmalar özellikle daha karmaşık veya dinamik hareket segmentlerinde belirginleşebilmekle birlikte, yöntemin genel yörünge takibi ve konum tahmini yeteneğini önemli ölçüde etkilememektedir. Orta zorluk seviyesindeki bu görsel sonuçlar önerilen yöntemin sensör gürültüsü, kısmi tıkanıklıklar veya daha karmaşık hareket desenleri gibi orta düzeydeki bozucu etkenlere karşı gürbüzlüğünü ortaya koymaktadır.

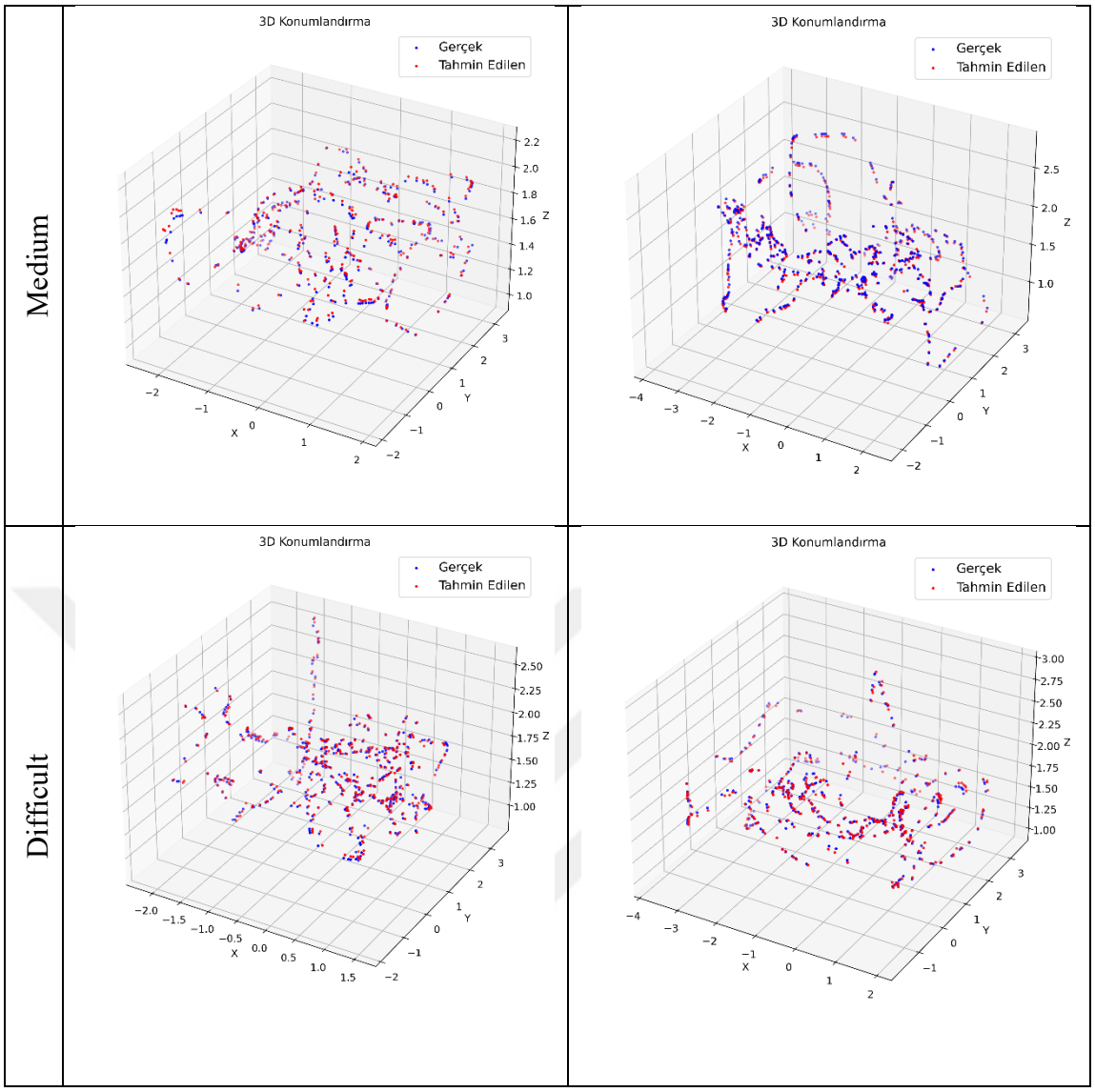
Zorlu zorluk seviyesindeki Vicon Room 1 ve Vicon Room 2 veri setleri incelendiğinde yöntemin hala gerçek yörüngeye oldukça yakın tahminler yapabildiği görülmektedir. Ancak özellikle yoğun dönüşler, hızlı hareketler veya görsel özelliklerin kısıtlı olduğu alanlarda kırmızı ve mavi noktalar arasındaki ayrımın bir miktar arttığı gözlemlenmektedir. Bu durum zorlu koşulların yöntemin performansını bir miktar etkileyebileceğini ancak yöntemin temel yörünge takibi ve konumlandırma kabiliyetini koruduğunu göstermektedir. Zorlu senaryolardaki bu kabul edilebilir performans önerilen

yöntemin dış ortam koşullarındaki değişkenliklere ve potansiyel sensör hatalarına karşı dikkate değer bir direnç sergilediğine işaret etmektedir. Genel olarak tüm zorluk seviyelerinde gerçek ve tahmin edilen yörüngeler arasındaki yakın uyum, önerilen yöntemin hem doğruluğu hem de farklı çevresel koşullara adaptasyon yeteneği açısından güçlü bir potansiyel sunduğunu doğrulamaktadır.

Tablo 4.7 3D konumlandırma grafikleri, önerilen görsel-atalet odometri (VIO) yönteminin Vicon Room 1 ve Vicon Room 2 veri setleri üzerindeki üç boyutlu yörünge tahmini yeteneğini farklı zorluk seviyelerinde detaylı bir şekilde gözler önüne sermektedir. Her bir grafikte mavi noktalar sistemin gerçek 3D konumunu ("Actual"), kırmızı noktalar ise önerilen yöntem tarafından tahmin edilen 3D yörüngeyi ("Predicted") temsil etmektedir. Bu görselleştirmeler, yöntemin üç boyutlu uzaydaki doğruluk, gürbüzlük ve genel performansına ilişkin derinlemesine bir bakış sağlamaktadır.

Tablo 4.7. 3D Konumlandırma Grafikleri

	Vicon Room 1	Vicon Room 2
Easy		



Kolay zorluk seviyesindeki Vicon Room 1 ve Vicon Room 2 veri setleri incelendiğinde, önerilen yöntemin gerçek yörünge ile başarılı bir örtüşme sergilediği açıkça görülmektedir. Tahmin edilen kırmızı noktaların gerçek mavi noktalarla sıkı bir şekilde çakışması yöntemin ideal veya nispeten az bozulmuş koşullar altında dahi son derece yüksek doğrulukta 3D konum ve yörünge tahmini yapabildiğini göstermektedir. Hem yatay düzlemdeki (XY) hareketler hem de dikey eksendeki (Z) değişimler gerçek verilerle başarılı bir şekilde eşleşmekte bu da yöntemin uzaysal farkındalığının ve hassasiyetinin altını çizmektedir.

Orta zorluk seviyesindeki Vicon Room 1 ve Vicon Room 2 veri setlerine geçildiğinde önerilen yöntemin 3D performansında da tutarlı bir devamlılık gözlemlenmektedir. Gerçek ve tahmin edilen yörüngeler arasındaki genel uyum hala yüksek seviyede olup yöntemin zorluk seviyesindeki artışa rağmen kararlı bir şekilde

çalıştığını göstermektedir. Yörünge dönüşlerinin veya yükseklik değişimlerinin yoğunlaştığı alanlarda kırmızı ve mavi noktalar arasında hafif ayrışmalar fark edilmiştir. Ancak bu sapmalar yörüngenin genel akışını veya 3D konumlandırma doğruluğunu temelden etkileyecek düzeyde değildir. Bu durum önerilen yöntemin orta dereceli sensör gürültüsü daha karmaşık hareket profilleri veya kısmi görsel belirsizlikler gibi 3D ortamdaki zorluklara karşı gürbüzlüğünü teyit etmektedir.

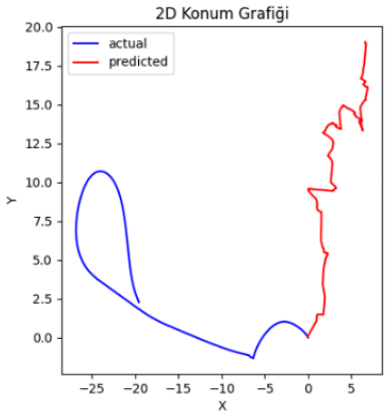
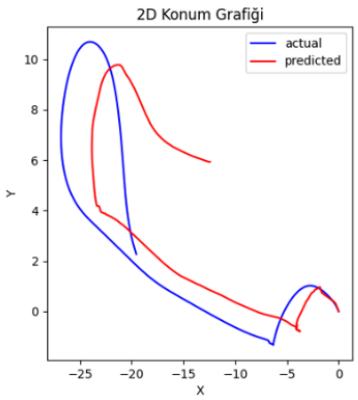
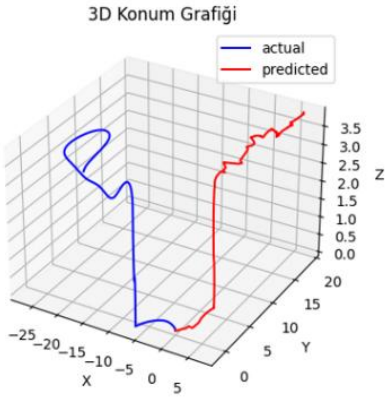
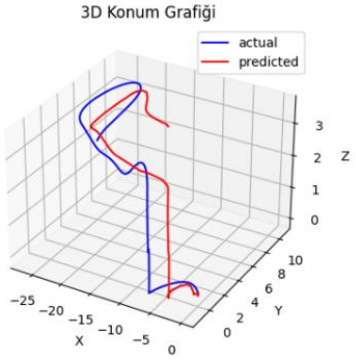
Zorluk seviyesi en yüksek olan Vicon Room 1 ve Vicon Room 2 veri setleri ile önerilen yöntemin sınırlarını test edilmiştir. Bu koşullar altında da tahmin edilen yörünge genel olarak gerçek yörüngeyi takip etmeyi başarmıştır. Yörünge boyunca özellikle hızlı hareketler, belirgin yükseklik değişiklikleri veya düşük görsel özellikli bölgelerde tahmin edilen ve gerçek konumlar arasında daha belirgin farklar ortaya çıkabilmektedir. Buna rağmen yöntemin 3D yörüngenin topolojisini ve genel eğilimini koruduğu ve geniş çaplı sapmalardan kaçındığı görülmektedir. Bu bulgu önerilen yöntemin zorlu ve gerçekçi ortamlarda dahi 3D konumlandırma görevleri için geçerli ve güvenilir sonuçlar üretebilme yeteneğine sahip olduğunu güçlü bir şekilde ortaya koymaktadır.

Sonuç olarak sunulan 3D konumlandırma grafikler, önerilen yöntemin hem kolay hem de zorlu senaryolarda yüksek doğruluk ve gürbüzlüğünü sergilediğini açıkça göstermektedir. Yöntemin üç boyutlu uzayda gerçek yörüngeyi etkin bir şekilde takip edebilmesi robotik navigasyon ve insansız hava araçları için önemli bir avantaj sağlamaktadır.

4.4. Simülasyon Pusula Sensörü Tam Konumlandırma Deneyi Sonuçları

Euroc MAV veri setinde sadece ataletsel veriler ile çalışan konumlandırma yöntemi başarısız olmuştur. Bu başarısızlığın EuRoc MAV veri setinde yalnızca jiroskop ve ivmeölçer içeren bir IMU kullanılmış olmasından kaynaklanıp kaynaklanmadığı araştırılmıştır. Bu araştırma için simülasyon ortamında bulunan İHA'dan yalnızca jiroskop ve ivmeölçer verileri toplanarak bir eğitim yapılmıştır. Ayrıca jiroskop, ivmeölçer ve pusulanın bulunduğu bir eğitim daha yapılmıştır. Yapılan eğitimler ile elde edilen modeller tüm sensörlerin bulunduğu eğitilmiş modeller ile bir uçuş üzerinde test edilmiş ve konumlandırma çıktıları alınmıştır. Alınan çıktıların karşılaştırılması Tablo 4.8'de verilmiştir.

Tablo 4.8. Farklı veri girdileriyle eğitilen model çıktıları

Jiroskop ve İvmeölçer Verilerinden Oluşan Eğitim	**Jiroskop, İvmeölçer ve Pusula Verilerinden Oluşan Eğitim**
	
	

Tablo 4.7'de elde edilen çıktılar incelendiğinde sadece jiroskop ve ivmeölçer içeren simülasyon verisi tam konumlandırma modeli çıktısının özellik çıkarma modelinin konumlandırma çıktılarına benzer şekilde eksik olduğu görülmüştür. Ancak jiroskop ve ivmeölçer sensörünün yanına pusula sensörünün eklenmesi ile eğitilen modelin konumlandırma çıktılarının tüm sensörleri içeren modelin konumlandırma çıktılarına yakın olmuştur. Yapılan bu test ile İMU verileri yanına pusula gibi basit ve ucuz sensörlerin eklenmesi ile tam konumlandırma yapılabileceği sonucuna varılmıştır.

5. SONUÇLAR VE ÖNERİLER

5.1 Sonuçlar

Bu çalışmada insansız hava araçları (İHA) için derin öğrenme tabanlı ataletsel konumlandırma sistemleri ve derin dikkat tabanlı görsel ataletsel konumlandırma sistemleri üzerine kapsamlı deneysel çalışmalar gerçekleştirilmiştir. EuRoc MAV ve Webots simülasyon veri setleri kullanılarak farklı derin öğrenme modelleri eğitilmiş ve performansları çeşitli metrikler (RMSE, MSE, MAE, R^2, MAPE) üzerinden değerlendirilmiştir.

İlk olarak, derin öğrenme tabanlı ataletsel konumlandırma sistemleri için yapılan testlerde dropout kullanımının genel olarak aşırı öğrenmeyi azalttığı ancak bazı modellerde hataları artırdığı gözlemlenmiştir. Elde edilen sonuçlara göre dropout içeren AHLSTM modeli tüm metriklerde en iyi performansı göstererek bu alandaki etkinliğini kanıtlamıştır. Bu modelin eğitim verisiyle birlikte daha önce görmediği yeni veriler üzerinde de başarılı konumlandırma yeteneğine sahip olduğu doğrulanmıştır. Ayrıca sistemin zaman içerisinde oluşan hatasının incelenmesiyle, hata grafiğinin sürekli artmak yerine zamanla dalgalandığı ve İHA'nın yükselme/alçalma gibi hareket durumlarına göre girdi değişkenliklerinden etkilendiği tespit edilmiştir. Bu durum, tam konumlandırıcı sistemin başarısının eksenlerdeki hareketlere göre değişebileceğini ortaya koymuştur.

İkinci olarak derin dikkat tabanlı görsel ataletsel konumlandırma sisteminin EuRoc MAV veri seti üzerindeki performansları, literatürdeki diğer çalışmalarla kıyaslanmıştır. Eğitim ve test kaybı eğrileri, farklı zorluk seviyeleri (Easy, Medium, Difficult) ve ortamlar (Vicon Room 1, Vicon Room 2) dikkate alınarak analiz edilmiştir. Kolay ve orta zorluk seviyelerinde modelin hızlı ve istikrarlı bir şekilde öğrendiği, ancak zorlu seviyelerde (özellikle Vicon Room 2'de) kayıplarda dalgalanmaların arttığı görülmüştür. Ortamlar arası karşılaştırmada, Vicon Room 1'deki sonuçların daha istikrarlı olduğu ve Vicon Room 2'nin daha fazla gürültü içerebileceği sonucuna varılmıştır. Genel olarak, modelin performansının hem görev zorluğuna hem de deney ortamının özelliklerine duyarlı olduğu ortaya konmuştur.

Literatür karşılaştırmasında, önerilen yöntem, RMSE metriği bazında altı farklı test veri setinde (V101, V102, V103, V201, V202, V203) üstün bir başarı sergilemiştir. Özellikle V101 ve V201 gibi kolay veri setlerinde rakiplerine kıyasla %50'nin üzerinde performans iyileştirmeleri sağlanmıştır. Ortalama RMSE değeri 0.032 ile karşılaştırılan

yöntemler arasında en iyi performans elde edilmiştir. Sıralama bazlı karşılaştırmada da önerilen yöntem 2.000 ortalama sıralama skoruyla tüm yöntemler arasında birinci sırada yer alarak farklı veri setlerinde tutarlı bir şekilde üstün performans sergilediğini kanıtlamıştır. 2D ve 3D konumlandırma grafikleri, önerilen yöntemin hem kolay hem de zorlu senaryolarda gerçek yörüngeye oldukça yakın tahminler yapabildiğini ve yüksek doğruluk ile gürbüzlük sunduğunu göstermiştir.

Son olarak simülasyon verisi ile yapılan ek bir deneyde, EuRoc MAV veri setinde sadece jiroskop ve ivmeölçer ile konumlandırmanın yetersiz kaldığı ancak pusula sensörünün eklenmesiyle konumlandırmanın mümkün olduğu tespit edilmiştir. Bu sonuç İMU verilerine basit ve ucuz sensörlerin entegrasyonuyla konumlandırma başarısının artırılabileceğini göstermiştir.

5.2 Öneriler

Bu çalışmanın bulguları ışığında gelecekteki araştırmalar için çeşitli yöntemler önerilebilir. İlk olarak modelin genelleme yeteneğini artırmak kritik öneme sahiptir. Bu özellikle zorlu ve değişken ortam koşullarında daha fazla veri çeşitliliği ve miktarı kullanarak eğitimi kapsayabilir; gürültülü veya kısmi tıkanıklıklar içeren veri setleri ekleyerek gerçekleştirilebilir. Ayrıca modelin aşırı öğrenmesini engellemek ve farklı senaryolara daha iyi uyum sağlaması için gelişmiş düzenleme (regularization) tekniklerinin araştırılması ve uygulanması faydalı olacaktır.

İkinci olarak sistemin gerçek dünya uygulamalarındaki performansını daha da yükseltmek amacıyla daha dinamik model mimarileri geliştirilebilir. Bu özellikle hareketli veya karmaşık çevrelerde adaptif öğrenme yeteneklerine sahip kendini ayarlayabilen ağ tasarımlarını içerebilir. Ayrıca konumlandırma hatasının zamana bağlı dalgalanmaları göz önüne alındığında bu hataları modelin öğrenme sürecine entegre eden veya öngörerek telafi eden hata modelleme mekanizmaları (örneğin, filtreleme yaklaşımları) entegre edilebilir. Gerçek zamanlı İHA uygulamaları için, modelin hesaplama verimliliğini artırmak adına model sıkıştırma veya daha optimize donanım platformlarına yönelik çalışmalar önem taşımaktadır.

Üçüncü olarak multi-modal sensör füzyonu yaklaşımları yalnızca görsel ve ataletsel verilerin ötesine geçerek, GPS, manyetometre (pusula), derinlik sensörleri veya Wi-Fi sinyalleri gibi ek sensör verilerini derin öğrenme tabanlı bir çerçevede birleştirmeyi hedefleyebilir. Bu, özellikle GPS sinyalinin zayıf olduğu veya görsel özelliklerin yetersiz

kaldığı durumlarda konumlandırma doğruluğunu ve gürbüzlüğünü önemli ölçüde artırabilir. Ayrıca CNN tabanlı yöntemlerin öğrendiği ortamda iyi çalışıp farklı ortamlarda aynı başarıyı gösterememe sorununa çözüm olarak daha genellenebilir görsel veri işleme tekniklerinin geliştirilmesi önemlidir. Bu özellik çıkarma aşamasında ortamdan bağımsız gürbüz temsiller öğrenmeyi hedefleyen yaklaşımları içerebilir veya transfer öğrenme ve alan adaptasyonu teknikleri ile modelin yeni ve bilinmeyen ortamlara daha hızlı adapte olması sağlanabilir. Bu sayede görsel konumlandırma sistemlerinin farklı, daha çeşitli gerçek dünya koşullarında güvenilir bir şekilde çalışabilmesi mümkün olacaktır.

KAYNAKLAR

Ahmad, Norhafizan, Raja Ariffin Raja Ghazilla, Nazirah M Khairi, ve Vijayabaskar Kasi. 2013. 'Reviews on various inertial measurement unit (IMU) sensor applications', *International Journal of Signal Processing Systems*, 1: 256-62.

Aslan, Muhammet Fatih, Akif Durdu, ve Kadir Sabanci. 2022. 'Visual-Inertial Image-Odometry Network (VIIONet): A Gaussian process regression-based deep architecture proposal for UAV pose estimation', *Measurement*, 194: 111030.

Aslan, Muhammet Fatih, Akif Durdu, Abdullah Yusefi, ve Alper Yilmaz. 2022. 'HVIOnet: A deep learning based hybrid visual–inertial odometry approach for unmanned aerial system position estimation', *Neural Networks*, 155: 461-74.

Bachrach, Abraham, Samuel Prentice, Ruijie He, ve Nicholas Roy. 2011. 'RANGE–Robust autonomous navigation in GPS-denied environments', *Journal of field robotics*, 28: 644-66.

Bai, Yu-ting, Zhi-yao Zhao, Xiao-yi Wang, Xue-bo Jin, ve Bai-hai Zhang. 2022. 'Continuous positioning with recurrent auto-regressive neural network for unmanned surface vehicles in GPS outages', *Neural Processing Letters*, 54: 1413-34.

Bai, Yuhao, Baohua Zhang, Naimin Xu, Jun Zhou, Jiayou Shi, ve Zhihua Diao. 2023. 'Vision-based navigation and guidance for agricultural autonomous vehicles and robots: A review', *Computers and Electronics in Agriculture*, 205: 107584.

Baldini, Francesca, Animashree Anandkumar, ve Richard M Murray. 2020. "Learning pose estimation for UAV autonomous navigation and landing using visual-inertial sensor data." In *2020 American Control Conference (ACC)*, 2961-66. IEEE.

Burri, Michael, Janosch Nikolic, Pascal Gohl, Thomas Schneider, Joern Rehder, Sammy Omari, Markus W Achtelik, ve Roland Siegwart. 2016. 'The EuRoC micro aerial vehicle datasets', *The International journal of robotics research*, 35: 1157-63.

Campos, Carlos, Richard Elvira, Juan J Gómez Rodríguez, José MM Montiel, ve Juan D Tardós. 2021. 'Orb-slam3: An accurate open-source library for visual, visual–inertial, and multimap slam', *IEEE transactions on robotics*, 37: 1874-90.

Caron, Francois, Emmanuel Duflos, Denis Pomorski, ve Philippe Vanheeghe. 2006. 'GPS/IMU data fusion using multisensor Kalman filtering: introduction of contextual aspects', *Information fusion*, 7: 221-30.

Chang, Yingxiu, Yongqiang Cheng, Umar Manzoor, ve John Murray. 2023. 'A review of UAV autonomous navigation in GPS-denied environments', *Robotics and Autonomous Systems*: 104533.

Clark, Ronald, Sen Wang, Hongkai Wen, Andrew Markham, ve Niki Trigoni. 2017. "Vinet: Visual-inertial odometry as a sequence-to-sequence learning problem." In *Proceedings of the AAAI conference on artificial intelligence*.

Couturier, Andy, ve Moulay A Akhloufi. 2021. 'A review on absolute visual localization for UAV', *Robotics and Autonomous Systems*, 135: 103666.

Dellaert, Frank. 2012. 'Factor graphs and GTSAM: A hands-on introduction', *Georgia Institute of Technology, Tech. Rep*, 2: 4.

Du, Shuang, Wei Sun, ve Yang Gao. 2017. 'Improving observability of an inertial system by rotary motions of an IMU', *Sensors*, 17: 698.

Faiçal, Bruno S, Heitor Freitas, Pedro H Gomes, Leandro Y Mano, Gustavo Pessin, André CPLF de Carvalho, Bhaskar Krishnamachari, ve Jó Ueyama. 2017. 'An adaptive approach for UAV-based pesticide spraying in dynamic environments', *Computers and Electronics in Agriculture*, 138: 210-23.

Fu, Qiang, Jialong Wang, Hongshan Yu, Islam Ali, Feng Guo, Yijia He, ve Hong Zhang. 2020. 'PL-VINS: Real-time monocular visual-inertial SLAM with point and line features', *arXiv preprint arXiv:2009.07462*.

Gálvez-López, Dorian, ve Juan D Tardos. 2012. 'Bags of binary words for fast place recognition in image sequences', *IEEE transactions on robotics*, 28: 1188-97.

Goncalves, J A, ve Renato Henriques. 2015. 'UAV photogrammetry for topographic monitoring of coastal areas', *ISPRS journal of Photogrammetry and Remote Sensing*, 104: 101-11.

Gupta, Abhishek, ve Xavier Fernando. 2022. 'Simultaneous localization and mapping (slam) and data fusion in unmanned aerial vehicles: Recent advances and challenges', *Drones*, 6: 85.

Gyagenda, Nasser, Jasper V Hatilima, Hubert Roth, ve Vadim Zhmud. 2022. 'A review of GNSS-independent UAV navigation techniques', *Robotics and Autonomous Systems*, 152: 104069.

Hadi, Ghozali Suhariyanto, Rivaldy Varianto, Bambang Riyanto Trilaksono, ve Agus Budiyono. 2014. 'Autonomous UAV system development for payload dropping mission', *Journal of Instrumentation, Automation and Systems*, 1: 72-77.

He, Ming, Chaozheng Zhu, Qian Huang, Baosen Ren, ve Jintao Liu. 2020. 'A review of monocular visual odometry', *The Visual Computer*, 36: 1053-65.

Hochreiter, Sepp, ve Jürgen Schmidhuber. 1997. 'Long short-term memory', *Neural computation*, 9: 1735-80.

Jiang, Changhui, Shuai Chen, Yuwei Chen, Boya Zhang, Ziyi Feng, Hui Zhou, ve Yuming Bo. 2018. 'A MEMS IMU de-noising method using long short term memory recurrent neural networks (LSTM-RNN)', *Sensors*, 18: 3470.

Kao, Peng-Yuan, Hsiu-Jui Chang, Kuan-Wei Tseng, Timothy Chen, He-Lin Luo, ve Yi-Ping Hung. 2023. 'VIUNet: deep visual–inertial–UWB fusion for indoor UAV localization', *Ieee Access*, 11: 61525-34.

Kendall, Alex, Matthew Grimes, ve Roberto Cipolla. 2015. "Posenet: A convolutional network for real-time 6-dof camera relocalization." In *Proceedings of the IEEE international conference on computer vision*, 2938-46.

Kerns, Andrew J, Daniel P Shepard, Jahshan A Bhatti, ve Todd E Humphreys. 2014. 'Unmanned aircraft capture and control via GPS spoofing', *Journal of field robotics*, 31: 617-36.

Leutenegger, Stefan. 2022. 'Okvis2: Realtime scalable visual-inertial slam with loop closure', *arXiv preprint arXiv:2202.09199*.

Leutenegger, Stefan, Simon Lynen, Michael Bosse, Roland Siegwart, ve Paul Furgale. 2015. 'Keyframe-based visual–inertial odometry using nonlinear optimization', *The International Journal of Robotics Research*, 34: 314-34.

Li, Bin, Zesong Fei, ve Yan Zhang. 2018. 'UAV communications for 5G and beyond: Recent advances and future trends', *IEEE Internet of Things Journal*, 6: 2241-63.

Li, Jian, Gongliu Yang, Qingzhong Cai, Haofei Niu, ve Jing Li. 2022. 'Cooperative navigation for UAVs in GNSS-denied area based on optimized belief propagation', *Measurement*, 192: 110797.

Li, Shunkai, Xin Wu, Yingdian Cao, ve Hongbin Zha. 2021. "Generalizing to the open world: Deep visual odometry with online adaptation." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 13184-93.

Lipson, Lahav, Zachary Teed, ve Jia Deng. 2024. 'Deep Patch Visual SLAM', *arXiv preprint arXiv:2408.01654*.

Menouar, Hamid, Ismail Guvenc, Kemal Akkaya, A Selcuk Uluagac, Abdullah Kadri, ve Adem Tuncer. 2017. 'UAV-enabled intelligent transportation systems for the

smart city: Applications and challenges', *IEEE Communications Magazine*, 55: 22-28.

Michel, Olivier. 2004. 'Cyberbotics ltd. webots™: professional mobile robot simulation', *International Journal of Advanced Robotic Systems*, 1: 5.

Mur-Artal, Raul, Jose Maria Martinez Montiel, ve Juan D Tardos. 2015. 'ORB-SLAM: a versatile and accurate monocular SLAM system', *IEEE transactions on robotics*, 31: 1147-63.

Mur-Artal, Raul, ve Juan D Tardós. 2017. 'Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras', *IEEE transactions on robotics*, 33: 1255-62.

Pan, Youqi, Wugen Zhou, Yingdian Cao, ve Hongbin Zha. 2024. "Adaptive vio: Deep visual-inertial odometry with online continual learning." In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 18019-28. IEEE.

Qin, Tong, Peiliang Li, ve Shaojie Shen. 2018. 'Vins-mono: A robust and versatile monocular visual-inertial state estimator', *IEEE transactions on robotics*, 34: 1004-20.

Rasmussen, Carl Edward. 2003. 'Gaussian processes in machine learning.' in, *Summer school on machine learning* (Springer).

Rosinol, Antoni, Marcus Abate, Yun Chang, ve Luca Carlone. 2020. "Kimera: an open-source library for real-time metric-semantic localization and mapping." In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 1689-96. IEEE.

Rublee, Ethan, Vincent Rabaud, Kurt Konolige, ve Gary Bradski. 2011. "ORB: An efficient alternative to SIFT or SURF." In *2011 International conference on computer vision*, 2564-71. Ieee.

Seiskari, Otto, Pekka Rantalankila, Juho Kannala, Jerry Ylilammi, Esa Rahtu, ve Arno Solin. 2022. "HybVIO: Pushing the limits of real-time visual-inertial odometry." In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 701-10.

Silva do Monte Lima, João Paulo, Hideaki Uchiyama, ve Rin-ichiro Taniguchi. 2019. 'End-to-end learning framework for imu-based 6-dof odometry', *Sensors*, 19: 3777.

Solodar, Dan, ve Itzik Klein. 2024. 'VIO-DualProNet: Visual-inertial odometry with learning based process noise covariance', *Engineering Applications of Artificial Intelligence*, 133: 108466.

Sudhakar, S, Varadarajan Vijayakumar, C Sathiya Kumar, V Priya, Logesh Ravi, ve V Subramaniyaswamy. 2020. 'Unmanned Aerial Vehicle (UAV) based Forest Fire Detection and monitoring for reducing false alarms in forest-fires', *Computer Communications*, 149: 1-16.

Taghizadeh, Sina, ve Reza Safabakhsh. 2023. 'An integrated INS/GNSS system with an attention-based hierarchical LSTM during GNSS outage', *GPS Solutions*, 27: 71.

Teed, Zachary, ve Jia Deng. 2021. 'Droid-slam: Deep visual slam for monocular, stereo, and rgb-d cameras', *Advances in neural information processing systems*, 34: 16558-69.

Teed, Zachary, Lahav Lipson, ve Jia Deng. 2024. 'Deep patch visual odometry', *Advances in neural information processing systems*, 36.

Tomic, Teodor, Korbinian Schmid, Philipp Lutz, Andreas Domel, Michael Kassecker, Elmar Mair, Iris Lynne Grixa, Felix Ruess, Michael Suppa, ve Darius Burschka. 2012. 'Toward a fully autonomous UAV: Research platform for indoor and

outdoor urban search and rescue', *IEEE robotics & automation magazine*, 19: 46-56.

Tu, Zheming, Changhao Chen, Xianfei Pan, Ruochen Liu, Jiarui Cui, ve Jun Mao. 2022. 'Ema-vio: Deep visual–inertial odometry with external memory attention', *IEEE Sensors Journal*, 22: 20877-85.

Valada, Abhinav, Noha Radwan, ve Wolfram Burgard. 2018. "Deep auxiliary learning for visual localization and odometry." In *2018 IEEE international conference on robotics and automation (ICRA)*, 6939-46. IEEE.

Von Gioi, Rafael Grompone, Jeremie Jakubowicz, Jean-Michel Morel, ve Gregory Randall. 2008. 'LSD: A fast line segment detector with a false detection control', *IEEE transactions on pattern analysis and machine intelligence*, 32: 722-32.

Wang, Sen, Ronald Clark, Hongkai Wen, ve Niki Trigoni. 2017. "Deepvo: Towards end-to-end visual odometry with deep recurrent convolutional neural networks." In *2017 IEEE international conference on robotics and automation (ICRA)*, 2043-50. IEEE.

Wen, Weisong, Tim Pfeifer, Xiwei Bai, ve Li-Ta Hsu. 2021. 'Factor graph optimization for GNSS/INS integration: A comparison with the extended Kalman filter', *NAVIGATION: Journal of the Institute of Navigation*, 68: 315-31.

Woo, Sanghyun, Jongchan Park, Joon-Young Lee, ve In So Kweon. 2018. "Cbam: Convolutional block attention module." In *Proceedings of the European conference on computer vision (ECCV)*, 3-19.

Yang, Mingyu, Yu Chen, ve Hun-Seok Kim. 2022. "Efficient deep visual and inertial odometry with adaptive visual modality selection." In *European Conference on Computer Vision*, 233-50. Springer.

Yuan, Huayu, Ke Han, ve Boyang Lou. 2024. 'Mix-VIO: A Visual Inertial Odometry Based on a Hybrid Tracking Strategy', *Sensors*, 24: 5218.

Zafari, Faheem, Athanasios Gkelias, ve Kin K Leung. 2019. 'A survey of indoor localization systems and technologies', *IEEE Communications Surveys & Tutorials*, 21: 2568-99.

Zheng, Tao, Aigong Xu, Xinchao Xu, ve Mingyue Liu. 2023. 'Modeling and compensation of inertial sensor errors in measurement systems', *Electronics*, 12: 2458.

Zhu, Ran, Mingkun Yang, Wang Liu, Rujun Song, Bo Yan, ve Zhuoling Xiao. 2022. 'DeepAVO: Efficient pose refining with feature distilling for deep Visual Odometry', *Neurocomputing*, 467: 22-35.

Zhuang, Licong, Xiaorong Zhong, Linjie Xu, Chunbao Tian, ve Wenshuai Yu. 2024. 'Visual SLAM for Unmanned Aerial Vehicles: Localization and Perception', *Sensors*, 24: 2980.

EKLER