

# TEKNOFEST Havacılık, Uzay ve Teknoloji Festivali Ulaşımında Yapay Zeka Yarışması Final Tasarım Raporu

FEZA  
320512

Üye 1: Burak AKAY  
Takım Kaptanı

Üye 2: Fatih YAŞAR  
İletişim Sorumlusu

Üye 3: Numan U. TÜRK  
Üye

Üye 4: Onur ALTUNTAŞ  
Üye

Üye 5: İsmail BAYRAKCI  
Üye

## Özet

Feza algoritması, drone veya uçan arabadan çekilmiş görüntülerden insan, taşıt, UAP (Uçan Araba Park) ve UAİ (Uçan Ambulans İniş) pistleri sınıflarını tespit etmek ve tespit edilen pistlerin inilebilir ya da inilemez olduğunun tespiti için tasarlanmıştır. Algoritma tasarımında insan ve taşıtların tespiti için Feza-base ve TPH-YOLOV5 [1] modelleri, pist tespiti için Feza-small modeli ve pist iniş durumu tespiti için PADIM (Patch Disribution Modeling) [2] kullanılmıştır. Tasarlanan modelin nesne tespit başarıları %43 AP, pist tespit başarıları %87 AP ve pist iniş durumu tespiti başarıları %99 F1 score olarak test verisi ile yapılan denemelerde ölçülmüştür. Feza algoritmasının hızı ise yaklaşık 1,18 it/s olduğu saptanmıştır.

## I. GİRİŞ

Günümüzde nesne tespit algoritmaları hızla gelişmektedir. Geliştirilen yeni modeller ile nesne tespit başarıları sürekli artmaktadır. Özellikle YOLO (You Only Look Once) [3] ve Swin [4] gibi algoritmalar gösterdikleri başarılar ile öne çıkmaktadır. Yeni modeller geliştirildikçe daha zor nesne tespit problemlerine çözüm aranmaya başlanmıştır. Bu problemlerin en önemlilerinden biri drone ile çekilen görüntülerdeki nesnelerin tespit edilmesidir. Drone görüntülerinde nesnelerin küçüklüğü, görüntü çözünürlüğünün düşük olması ve hava şartlarının farklı olması nesne tespit modelleri için büyük sıkıntılara yol açmaktadır. Feza algoritması bu problemlere çözüm arayarak hızlı, yüksek doğruluk oranına sahip ve görev odaklı olarak drone görüntülerinde nesne ve pist tespiti yapmak için geliştirilmiştir. Literatürde problemin çözümünde kullanmak üzere DBNet, SOLOER ve Cascade++ [5] gibi algoritmalar mevcuttur.

DBNet, Cascade R-CNN metodunu kullanmaktadır. Daha doğru sonuçlar üretebilmek için deforme edilebilir konvolüsyon blokları tercih edilmiştir. Backbone olarak ResNext-101 modeli tercih edilmiştir. Randomflip, Cutout ve Gaussian noise gibi çeşitli veri çoğaltma aşamalarıyla algoritmanın genelleme performansı artırılmıştır. Ulaşımda Yapay Zeka yarışmasına çok benzeyen Visdrone 2021 yarışmasında %65.34 AP50 başarılarıyla birinci olmuştur.

SOLOER, transformer tabanlı self-attention mekanizmalarını kullanmıştır. Metod olarak scaled-YOLOV4 modelini kullanmıştır. Küçük nesnelerin tespitini kolaylaştırmak için neck olarak BiFPN (Bidirectional Feature Pyramid Network) modelini tercih edilmiştir. Visdrone verileri ile yapılan testlerde %63.91 AP50 skorunu elde etmişlerdir.

Cascade++ ise metod olarak Cascade RCNN algoritmasını kullanmıştır. Anchor-free olmasıyla anchor kutularının sınırlamalarından kendini arındırarak farklı şekillerdeki nesnelerin tespitini kolaylaştırmıştır. Visdrone test veri setiyle yapılan değerlendirmede %62.92 AP50 skorunu elde ederek başarı sağlamıştır.

Geliştirilen Feza algoritması ise, bahsedilen algoritmalarından farklı olarak Ulaşımda Yapay Zeka yarışmasının beklentilerini karşılayacak şekilde en güncel çözüm yöntemlerini bir araya getirmiş ve literatüre yeni bakış açıları kazandırmıştır. Ulaşımda Yapay Zeka problemine en doğru çözümü getirmek için nesne ve pist tespitleri birbirinden ayrılmıştır. Nesne tespitinde kullanmak

üzere Backbone olarak ConvNeXt [6], neck için FPN (Feature Pyramid Network) [7] ve head algoritması için TOOD (Task-aligned One-stage Object Detection) [8] metodunu kullanmıştır. Anchor-free olması nedeniyle özellikle sürekli şekilleri değişen insan nesnesinin tespitini kolaylaştırmıştır. Genelleme performansını artırmak için Ulaşımda Yapay Zeka yarışmasının sağladığı veriler dışında Visdrone [5], UAV (Unmanned Aerial Vehicle) [9], Aerial dataset [10] ve AFO (Aerial Floating Objects) [11] gibi verilerden yararlanılmıştır. Doğruluğu artırmak için WBF (Weighted Boxes Fusion) [12] yöntemini kullanarak ensemble model oluşturulmuştur. Böylece Visdrone veri setiyle eğitilen ve açık kaynaklı olan TPH-YOLOV5 modelinin avantajlı yönleri Feza algoritmasına eklenmiştir. Bahsedilen diğer algoritmalarından farklı olarak Ulaşımda Yapay Zeka verisi iniş pistlerini de içermektedir. İniş pistlerinin tespiti için Feza-Base modelinin daha az parametrelili versiyonu kullanılmıştır. Böylece nesne ve pistlerin tespiti birbirinden ayrılarak görev odaklı bir algoritma hedeflenmiştir. Buna ek olarak problemin bir parçası olan iniş durum tespiti için PADIM anomali tespit modeli kullanılmıştır. Gaus dağılımına [13] göre iniş durum tespiti yapan model, Feza algoritmasına eklenerek Ulaşımda Yapay Zeka yarışmasında kullanmak üzere tercih edilmiştir. Literatürde bulunan birçok modelden daha güncel ve Ulaşımda Yapay Zeka'nın getirmiş olduğu pist tespiti ve pist iniş durumu tespiti ile ekstra özelliklere sahip olan Feza algoritması pist iniş durumu tespiti için anomali tespit sistemlerini kullanmak ve yeni nesne tespit modelleri üretmek gibi literatüre yeni yaklaşımlar kazandırmıştır.

## II. KULLANILAN VERİ SETLERİ

Model başarısını en çok etkileyen faktörlerden biri veri setinin kalitesidir. Veri seti probleme uygun olarak seçilmelidir. Problem, drone görüntülerindeki insan, taşıt ve pist gibi nesnelerin tespiti olduğu için bu kapsamda veri setleri seçilmiştir.

Visdrone veri seti, problemin çözümünde kullanılmıştır. Bu veri setinin tamamı insan ve taşıt nesnelerinin çekildiği drone görüntülerinden oluşur. Yaklaşık 10209 görüntüye sahiptir ve bu görüntüler farklı yüksekliklerde ve açılarda çekilmiştir. Görüntülerin neredeyse tamamı etiketlidir. Ulaşımda Yapay Zeka problemine çok uygun bir veri setidir. Feza-Base nesne tespit algoritması eğitilirken kullanıldı. Bununla modelin genelleme performansı artırılmak istendi.

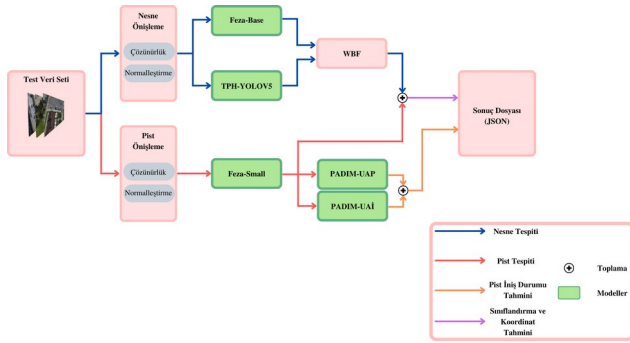
Diğer kullanılan veri seti UAV'dır. Tamamı drone görüntülerinden oluşan bu veri seti, trafikteki araçların tespit problemi üzerine oluşturulmuştur. Yaklaşık 80000 etiketli görüntülerden oluşur. Visdrone'dan farklı olarak tamamen araçlara odaklanmıştır ve yağmurlu, güneşli ve gece gibi hava koşulları düşünülerek hazırlanmıştır. UAV, eğitim veri setinin çeşitliliğini artırmak ve modelin farklı araç tiplerine ve hava koşullarına hazır olmasını sağlamak amacıyla kullanılmıştır.

Aerial Dataset, Kaggle sitesinden alınmış açık kaynaklı bir veri setidir. Drone görüntülerinden oluşur ve insanların tespitine odaklanmıştır. Yaklaşık 1200 görüntüden oluşur ve görüntüler insanların farklı şekillerde olduğu parklarda çekilmiştir. Oturan, spor yapan, yürüyen ve koşan insanların

bulunduğu bu veri seti modelin insanları tanımasını kolaylaştırmak için seçilmiş ve kullanılmıştır. AFO, yine Kaggle sitesinden alınmış açık kaynaklı bir veri setidir. Plajdaki insanların ve deniz araçlarının tespitine odaklanmış bir veri setidir. Yaklaşık 3647 etiketli görüntülerden oluşur. Tüm görüntüler Ulaşımda Yapay Zeka'ya uygun bir şekilde drone ile çekilmiştir. İnsanlar yüzerken, yatarken ve eğlenirken çekilmiş görüntülerdir. Ayrıca denizde bulunan taşıtları içermesiyle çeşitliliği artırmaktadır. Pist tespitinde kullanmak üzere, yarışma komitesinin sağladığı pist görüntülerinden yararlanılmıştır. UAP ve UAİ görüntüleri çeşitli görüntülere eklenerek yeni bir veri seti oluşturulmuştur. Bu şekilde yalancı etiketleme yapılmıştır. Buna ek olarak yarışmanın paylaştığı videolardan alınan pist içeren görüntüler de kullanılmıştır. Yine de pist veri sayısı istenilen düzeye gelmediği için döndürme, canlılık, parlaklık ve bulanıklık gibi çeşitli veri çoğaltma adımları uygulanarak pist verileri çoğaltılmıştır. Veriler çoğaltılmadan önce 541 görüntüden oluşuyordu. Çoğaltma işleminden sonra yaklaşık 1381 görüntüye ulaşıldı. Toplam UAP ve UAİ nesne sayılarının eşit olmasına dikkat edildi. Böylece eğitilen modelin bir pist çeşidine odaklanması önlenmiştir.

### III. YÖNTEM

#### A. Model Mimarisi



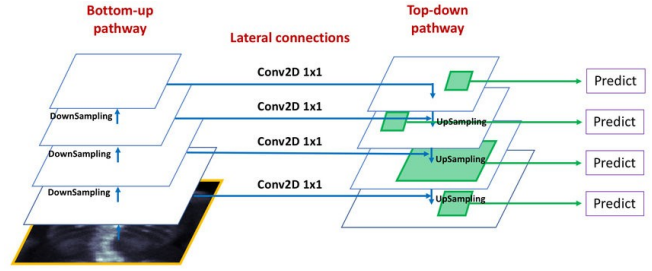
Figür 1: Feza Algoritma mimarisi

Ulaşımda Yapay Zeka problemine çözüm olarak kullanılan algoritmalar tasarlanırken özgünlüğe ve görev odaklı olmalarına özen gösterildi. Problemin çözümünde kullanılacak algoritma 3 ana parçadan oluşuyor; bunlar nesne tespiti, pist tespiti ve pist iniş durumu tespitidir. Tasarlanan algoritmanın bu şekilde parçalardan oluşmasındaki en büyük sebep, pist tespitine ayrı önem veriliyor olmasıdır. Pist ve nesnelerin tespiti farklı modeller ile gerçekleştirilerek algoritma başarısı artırıldı. Buna ek olarak görev odaklı modeller tasarlanmış oldu.

Nesne tespiti bölümündeki amaç insan ve taşıtların doğru bir şekilde sınıflandırılıp görüntüye göre koordinatlarının saptanmasıdır. Bu problemi çözmek için Feza-Base modeli geliştirildi. Backbone olarak ConvNeXt, neck olarak FPN ve head olarak t-head [8] kullanıldı.

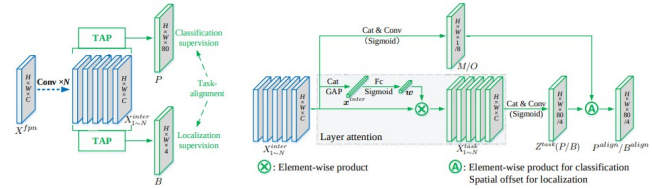
Backbone olarak tercih edilen ConvNext, çok güncel bir algoritmadır. Nesne tespitinde sıklıkla kullanılan ResNet [14] algoritmasının modernize edilmiş halidir. ResNet'in blok sayıları değiştirildi, depth-wise konvolüsyon blokları eklendi

ve darboğaz hücrelerinin yerine ters yönlü darboğaz hücreleri tercih edildi. Böylece ResNet geliştirilerek ConvNext algoritmasına dönüştürüldü. Yapılan araştırmalar sonucunda bu algoritma muadilleri arasında çok daha başarılı olduğu saptandı ve backbone olarak tercih edildi.



Figür 2: Feature Pyramid Network mimarisi

Birçok nesne tespit algoritmalarında neck olarak kullanılan FPN hafif, hızlı ve başarıyı artıran etken olması neticesinde Feza modelinde neck olarak kullanıldı. FPN, yukarıdan aşağıya ve aşağıdan yukarıya doğru katmanlardan oluşur ve öznetelik kaybını azaltır. Özellikle küçük nesnelerin tespitini kolaylaştırır. Bu avantajları sayesinde nesne tespit modelimize eklendi.



Figür 3: Soldaki T-head mimarisi ve sağdaki T-head mimarisinde kullanılan Task Aligned Predictor (TAP) bloğu

Nesne tespit modellerinin karar verici bölümü head algoritmalarıdır. Sınıflandırma ve koordinat tahminleri backbone ve neck'ten gelen verilere göre bu kısımda belirlenir. Modelde head olarak tercih edilen T-head, metod olarak TOOD yolunu izler; en önemli özellikleri anchor-free olması ve görev odaklı tasarlanmış olmasıdır. Tek aşamalı olan T-head ayrıca sınıflandırma ve koordinat çıktıları arasında ilişki kurar ve daha doğru sonuçlar üretir. Hem hız hem de performans açısından başarılı olduğu testlerde görülmüş ve Feza nesne tespit algoritmasına head olarak eklenmiştir.

TPH-YOLOV5, Feza-Base modelinin başarısını artırmak için Feza algoritmasına eklendi. Bu algoritma normal YOLOV5'den farklı olarak transformer blokları içerir ve Visdrone yarışmasında kullanılmak üzere geliştirilmiştir. Tamamen açık kaynaklıdır. Probleme uyumluluğu açısından ve gösterdiği başarılı performanstan dolayı nesne tespit algoritmasında kullanıldı. Feza-Base ve TPH-YOLOV5, WBF metoduyla birleştirildi ve güçlü, probleme uygun nesne tespit algoritması geliştirildi. WBF, her iki modelin çıktılarını analiz eder ve avantajlı yönlerini birleştirmeye olanak sağlar. Böylece Feza-Base'in açıklarını TPH-YOLOV5, TPH-YOLOV5'in açıklarını Feza-Base modeli kapatır.

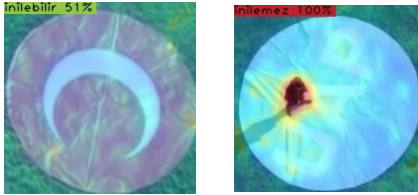
Pist tespiti için Feza-Base modelinin daha az parametrelili, daha hızlı ve daha az derin versiyonu kullanıldı. Nesne tespiti için tasarlanan Feza-Base modelinin backbonedaki her katmanı için derinlik sayıları (3, 3, 27, 3) olarak ayarlanmışken pist tespiti için oluşturulan Feza-Small

modelinin her katmanı için derinlik sayıları (3, 3, 9, 3) olarak belirlenmiştir. Aradaki bu farktan dolayı nesne tespiti için tasarlanan Feza modeline Feza-Base modeli, pist tespiti için tasarlanan versiyonuna Feza-Small modeli denmesi tercih edildi. Nesne tespiti ve pist tespiti modellerinin eğitimindeki bir diğer fark ise görüntü çözünürlüğüdür. Feza-Base 1280X466 çözünürlüklü görüntüler ile eğitilirken Feza-Small modeli için 640X360 boyutlu görüntüler kullanılmıştır. Bunlar dışında Feza-Base ve Feza-Small model özellikleri aynıdır. Bu ayrımın yapılmasındaki amaç, görev odaklı ve daha başarılı bir algoritma tasarlamaktır.

Model	Backbone	Neck	Head	Derinlik	Filtre	Metot	Hedef
Feza-Base	ConvNeXt	FPN	T-head	(3, 3, 27, 3)	(96, 192, 384, 768)	TOOD	Nesne Tespiti
Feza-Small	ConvNeXt	FPN	T-head	(3, 3, 9, 3)	(96, 192, 384, 768)	TOOD	Pist Tespiti
TPH-YOLOV5	CSPDarknet53	BiFPN	YOLO-head	(3, 6, 9, 3)	(128, 256, 512, 1024)	YOLO	Nesne Tespiti

Tablo 1: Nesne ve pist tespitinde kullanılan modellerin özellikleri

Pist iniş durumu tespiti içinse anomali tespit sistemleri kullanıldı. Anomali tespit sistemleri normalde seri üretimlerdeki ürünlerin normal olmayanlarını veya istenmeyen özelliklerde olanlarını tespit etmek için tasarlanan algoritmalar. Anomali tespit algoritmalarından PADIM, problem çözümünde kullanmak üzere seçildi ve çok başarılı oldu. PADIM'in en büyük avantajları çok az veri istiyor, çok hızlı eğitilebiliyor ve hızlı ve doğru sonuçlar alınabiliyor olmasıdır. PADIM, öznetelik çıkarımı için ResNet-18 algoritmasının önceden ImageNet ile eğitilmiş versiyonunu kullanıyor. Böylece kısa süren eğitimlerde bile doğru öznetelik çıkarımı yapabiliyor. Çıkarılan öznetelikler Gauss dağılımına göre hesaplanarak modele verilen nesnenin normal olduğu ya da anomali içerdiği tahmin ediliyor. PADIM, anormal (inilemez) pist görüntülerine gerek kalmadan sadece normal (inilebilir) pist görüntüleri ile eğitilebiliyor. Daha doğru sonuçlar elde edebilmek adına UAP ve UAI için 2 ayrı ama aynı özellikteki PADIM modelleri eğitildi.



Figür 4: PADIM algoritmasının pist iniş durumu tespitleri. Resimlerdeki kırmızı alanlar Gauss dağılımının yüksek olduğu bölgeleri gösterir.

#### B. Modelin Gerçekleştirme Adımları

Figür 1, algoritmamızın çalışma mantığıdır. Görüntüler nesne ve pist tespiti modellerine uygun hale getirilmek üzere önışlemden geçirilir. Piksel değerleri normalleştirilir ve çözünürlük, görüntünün verileceği modele göre ayarlanır. Nesne tespiti için görüntü Feza-Base ve TPH-YOLOV5 modellerinden geçirilir. Tahmin yüzdesi %48'den fazla ise çıktı kabul edilir. Her iki modelden çıkan tahminler WBF algoritmasıyla analiz edilerek belirlenen şartlarda birleştirilir. Birleşim sırasında her nesne için yeni bir tahmin yüzdesi hesaplanır. Eğer tahmin yüzdesi %30'un altındaysa o nesne yok sayılır. Nesne olmasa bile nesnemiş gibi etiket yapılması (False Positive) bu şekilde engellenir. Pist tespiti

için aynı görüntü Feza-Small modelinden geçirilir. Tahmin başarısı %87'nin üzerindeyse o koordinatların pist olduğu varsayılır. Eğer pist tespiti yapıldıysa görüntünün sadece pist koordinatlarını içeren kısmı, pist sınıfına uygun olarak PADIM-UAP veya PADIM-UAI modeline verilir. PADIM modeli pistlerin inilebilir veya inilemez olduğunu Gauss dağılımına göre tahmin eder. Eğer Gauss dağılımı %75'in üzerinde ise pist inilemezdir. Aksi halde pist inilebilir durumdadır. Feza-Small modelinden çıkan pist tahmini WBF çıktısıyla toplanır ve elde edilen sınıflandırma ve koordinat tahminleri ile pist iniş durumu tahminleri JSON halinde raporlanarak Ulaşımda Yapay Zeka formatına çevrilir.

#### IV. SONUÇLAR

Algoritma	bbox-mAP	bbox-mAP 50
EfficientDet	38.51	63.25
DNEFS	38.53	62.86
cascade++	38.72	62.92
VistrongerDet	38.77	64.28
Feza	39.18	62.83

Tablo 2: Literatürde bulunan ve drone görüntüleri için eğitilmiş bazı modellerin Visdrone 2021 object detection verileri üzerindeki AP cinsinden başarı sonuçları

Visdrone veri seti ile yapılan testlerde literatürde bulunan diğer algoritmalarla göre Feza algoritmasının başarısı yukarıdaki gibidir. Test verisi olarak Visdrone 2021 object detection challenge'dan [5] alınmış 3190 test görüntüsü kullanılmıştır. Bu görüntülerde insan ve taşıtlar mevcuttur. Ulaşımda Yapay Zeka verilerine benzer şekilde görüntüler drone ile çekilmiştir. Görüntüler gece ve gündüz gibi farklı hava koşullarını içermektedir. Visdrone test görüntüleriyle yapılan testlerde Feza algoritması AP ölçüm standartlarında muadili olan birçok modelden daha başarılı sonuç göstermiştir.

Model	Veri Seti	Hız (it/s)	bbox-mAP
Feza-Small	Pist Tespit	6.10	0.876

Tablo 3: Feza-small modelinin Ulaşımda Yapay Zeka 2022 test verisindeki pist tespit başarısı

Pist tespiti için kullanılan Feza-Small modelinin göstermiş olduğu başarı yukarıdaki tablodaki gibidir. Ulaşımda Yapay Zeka 2022 örnek videodan alınan framelemlerden elde edilen görüntüler kullanılarak gerçekleştirilmiş olan testlerde Feza-Small modeli %87 AP başarı elde etmiştir. Çok hızlı olması nedeniyle sisteme çok yük bindirmeyerek pist tespitini gerçekleştirebilir. Saniyede yaklaşık 6 görüntü işleyebilir.

Model	Veri Seti	Hız (it/s)	AUROC	F1
PADIM-UAI	Pist İnış Durum Tespit	6.78	0.999	0.995
PADIM-UAP	Pist İnış Durum Tespit	7.47	0.998	0.993

Tablo 4: PADIM modelinin Ulaşımda Yapay Zeka tarafından paylaşılan 2022 test videosundaki pist iniş durumu tespit başarısı. UAI pisti için PADIM-UAI, UAP pisti için PADIM-UAP modelleri kullanılmıştır.

Feza Algoritmasının pist iniş durumu tespit başarısı yukarıdaki gibidir. Pist iniş durumu tespiti için getirilen yeni yaklaşım olan anomali tespit mekanizmaları başarılı sonuçlar elde etmiştir. Ulaşımda Yapay Zeka komitesinin paylaşmış olduğu örnek videodaki framelelerden alınan veriler ile yapılan testlerde Feza algoritmasının pist iniş durum tespiti modülü olan PADIM yaklaşık 0.99 F1 skoru ile yüksek başarı sergilemiştir. Ayrıca çok hızlı olduğu için sisteme fazla yük bindirmemektedir. İki farklı iniş pisti bulunduğu için iki farklı ama aynı özellikteki modeller kullanılmıştır.

## V. DEĞERLENDİRME

Bu çalışmada drone ve uçan arabalar tarafından çekilmiş görüntülerde insan, taşıt ve pist nesnelerinin tespiti ve pistlerin inilebilir veya inilemez olduğunun tespiti için Feza algoritması geliştirilmiştir. Visdrone 2021 object detection verileriyle yapılan testlerde özellikle insan ve taşıt tespitinde Feza algoritması Cascade++, ViStrongerDet ve DNEFS gibi algoritmaları geride bırakmış ve daha başarılı sonuç göstermiştir. Nesne tespitinde kullanmak üzere yeni ve güncel bir model tasarlanmıştır. Bu model Feza-Base olarak isimlendirilmiştir. Özellikle drone görüntülerinde kullanmak üzere literatüre yeni bir nesne tespit algoritması kazandırılmıştır. UAP ve UAİ iniş pislerinin tespitinde geliştirilen yeni Feza-Small modeli ile yüksek başarılar elde edilmiştir. Literatürde anomali tespiti için tasarlanan PADIM modeli Feza algoritmasına eklenmiş ve pist iniş durum tespiti için kullanılmıştır. Pist iniş durumu tespitinde çok yüksek başarılar elde edilmiştir. Böylece pist iniş durumu tespiti için anomali tespit modeli kullanılarak literatüre yeni bir yaklaşım kazandırılmıştır.

Her ne kadar Feza algoritmasıyla yüksek başarılar elde edilmiş olsa da bazı sınırlamalar mevcuttur. Algoritmanın çok derin olması neticesinde hız olarak gerçek zamanlı değildir. Geliştirilen E-ELAN [3] gibi yeni yöntemler uygulanarak algoritma derinliği azaltılabilir. Ayrıca insanların 90 derece açındaki görüntüleri çok küçük olduğu için model, özellikle karlı havalarda bazı karlı bölgeleri karıştırıp insan olarak etiketlediği tespit edilmiştir. Tespiti yapılan insan koordinatlarının gaus dağılımı analizi yapılarak bahsedilen false positive durumların önüne geçilebilir.

Genel anlamda hem insan, taşıt ve pist nesnelerinin tespitini hem de pist iniş durumu tespitini yüksek doğruluk ve Ulaşımda yapay zeka kapsamında yeterli hızla yapabilen Feza algoritması literatüre kazandırdıklarıyla ve Ulaşımda Yapay Zeka ve Visdrone verileriyle yapılan testlerdeki başarılarıyla güncel algoritmalar arasında çok iyi sonuçlar elde etmiştir.

## REFERANSLAR

- [1] Zhu, X., Lyu, S., Wang, X., & Zhao, Q. (2021). TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 2778-2788).
- [2] Defard, T., Setkov, A., Loesch, A., & Audigier, R. (2021, January). Padim: a patch distribution modeling framework for anomaly detection and localization. In *International Conference on Pattern Recognition* (pp. 475-489). Springer, Cham.
- [3] Wang, C. Y., Bochkovskiy, A., & Liao, H. Y. M. (2022). YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. *arXiv preprint arXiv:2207.02696*.
- [4] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10012-10022).
- [5] Cao, Y., He, Z., Wang, L., Wang, W., Yuan, Y., Zhang, D., ... & Liu, M. (2021). VisDrone-DET2021: The vision meets drone object detection challenge results. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 2847-2854).
- [6] Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., & Xie, S. (2022). A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 11976-11986).
- [7] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125).
- [8] Feng, C., Zhong, Y., Gao, Y., Scott, M. R., & Huang, W. (2021, October). Tood: Task-aligned one-stage object detection. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (pp. 3490-3499). IEEE Computer Society.
- [9] Zeng, Y., Duan, Q., Chen, X., Peng, D., Mao, Y., & Yang, K. (2021). UAVData: A dataset for unmanned aerial vehicle detection. *Soft Computing*, 25(7), 5385-5393.
- [10] <https://www.kaggle.com/datasets/mysteryprsm/aerial-dataset>
- [11] <https://www.kaggle.com/datasets/jangsienicajzkowj/afo-aerial-dataset-of-floating-objects>
- [12] Solovyev, R., Wang, W., & Gabruseva, T. (2021). Weighted boxes fusion: Ensembling boxes from different object detection models. *Image and Vision Computing*, 107, 104117.
- [13] Goodman, N. R. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *The Annals of mathematical statistics*, 34(1), 152-177.
- [14] Targ, S., Almeida, D., & Lyman, K. (2016). Resnet in resnet: Generalizing residual architectures. *arXiv preprint arXiv:1603.08029*.