

# **CIFAR-10 Üzerinde 5 Model Karşılaştırması (Aynı Şartlar)**

## **CNN ve Vision Transformer Modellerinin Karşılaştırmalı Analizi**

### **Modeller: SimpleCNN, CustomCNN, ResNet-18, ViT-Tiny, ConvNeXt-Tiny**

Derin Öğrenme Uygulamaları Dersi Bütünleme

Fatma Nur Sakal

254217006 / Bilişim Sistemleri Mühendisliği Tezli Yüksek Lisans

Muğla Sıtkı Koçman Üniversitesi

[fatmanursakal@posta.mu.edu.tr](mailto:fatmanursakal@posta.mu.edu.tr)

# Motivasyon ve Problem Tanımı

Görüntü sınıflandırma problemlerinde mimari seçimi (CNN vs Transformer) model başarımını belirgin biçimde etkileyebilir. Ancak mimariler farklı giriş boyutu, farklı augmentation, farklı eğitim protokolü ile çalıştırıldığında karşılaştırma adil olmaz. Bu çalışma, CIFAR-10 üzerinde farklı mimarileri aynı şartlarda çalıştırarak “mimari etkisi”ni ayırtırmayı hedefler.

## Adil Karşılaştırma Sorunu

- Literatürde “en iyi model” söylemleri çoğu zaman farklı ön-isleme ve farklı eğitim ayarlarıyla raporlanır.
- Bu durum, performans farkının mimariden mi yoksa pipeline farklarından mı geldiğini belirsizleştirir

## CIFAR-10 Verisetinin Rolü

- CIFAR-10 küçük boyutlu ( $32 \times 32$ ) ama zengin sınıf içeriğine sahip klasik bir benchmark’tır.
- Kısa sürede çoklu model deneyi yapmak ve karşılaştırmalı analiz üretmek için uygundur.

## Otomatik ve Tekrarlanabilir Pipeline İhtiyacı

- Çoklu modeli tek tek eğitmek zaman alır ve hata riski oluşturur.
- Bu proje, tek komutla eğitim + değerlendirme + görselleştirme üreten bir repo akışı ile tekrarlanabilir sonuç üretir.

Hedef: CIFAR-10 üzerinde SimpleCNN, CustomCNN, ResNet-18, ViT-Tiny, ConvNeXt-Tiny performanslarını aynı protokol ile karşılaştırmak.

# Çalışmanın Amacı ve Araştırma Soruları

Bu çalışma, CIFAR-10 üzerinde farklı mimarilerin sınıflandırma performansını sistematik ve standardize bir protokol ile karşılaştırır. Uygulama geliştirmeden ziyade karşılaştırmalı analiz ve yorum odaklıdır.

## Model Karşılaştırması

- Aynı giriş boyutu ve benzer eğitim düzeni altında CNN tabanlı modeller mi, yoksa Transformer/modern mimariler mi daha başarılı?
- Mimari karmaşıklığı (ResNet/ConvNeXt/ViT) ile doğruluk artışı arasında gözlenen ilişki nedir?

## Veri Artırma (Augmentation) Etkisi

- Aynı model için augmentation açık/kapalı senaryosu performansı anlamlı ölçüde değiştirmiyor mu?
- Özellikle ResNet-18 üzerinde “Aug ON vs OFF” ablation sonucu ne söylüyor?

## Kritik Metrik ve Güvenilirlik

- Bu projede ana metrik accuracy (val/test) olarak raporlanmıştır.
- Ek olarak confusion matrix ve sınıf bazlı hataların analizi ile “hangi sınıflar karışıyor?” sorusu yanıtlanmıştır.

❑ Önemli Not: Bu çalışma bir “yeni model icadı” değil; aynı şartlarda çoklu mimari karşılaştırması ve görsel/istatistiksel analiz projesidir.

# CIFAR-10 BAĞLAMI

CIFAR-10, 10 sınıfı doğal görüntüler içeren ve bilgisayarlı görü alanında yaygın kullanılan bir benchmark veri setidir. Bu veri seti, düşük çözünürlüklü görüntüler ( $32 \times 32$ ) ile çalıştığı için hızlı prototipleme sağlar; ancak modern mimariler çoğunlukla  $224 \times 224$  giriş beklediğinden, bu projede adil karşılaştırma amacıyla tüm modeller  $224 \times 224$  girişe standardize edilmiştir.

**Vurgu:** Bu standardizasyon, mimari karşılaştırmasını “farklı input boyutu etkisi”nden arındırmayı amaçlar.

# Veri Seti Tanıtımı

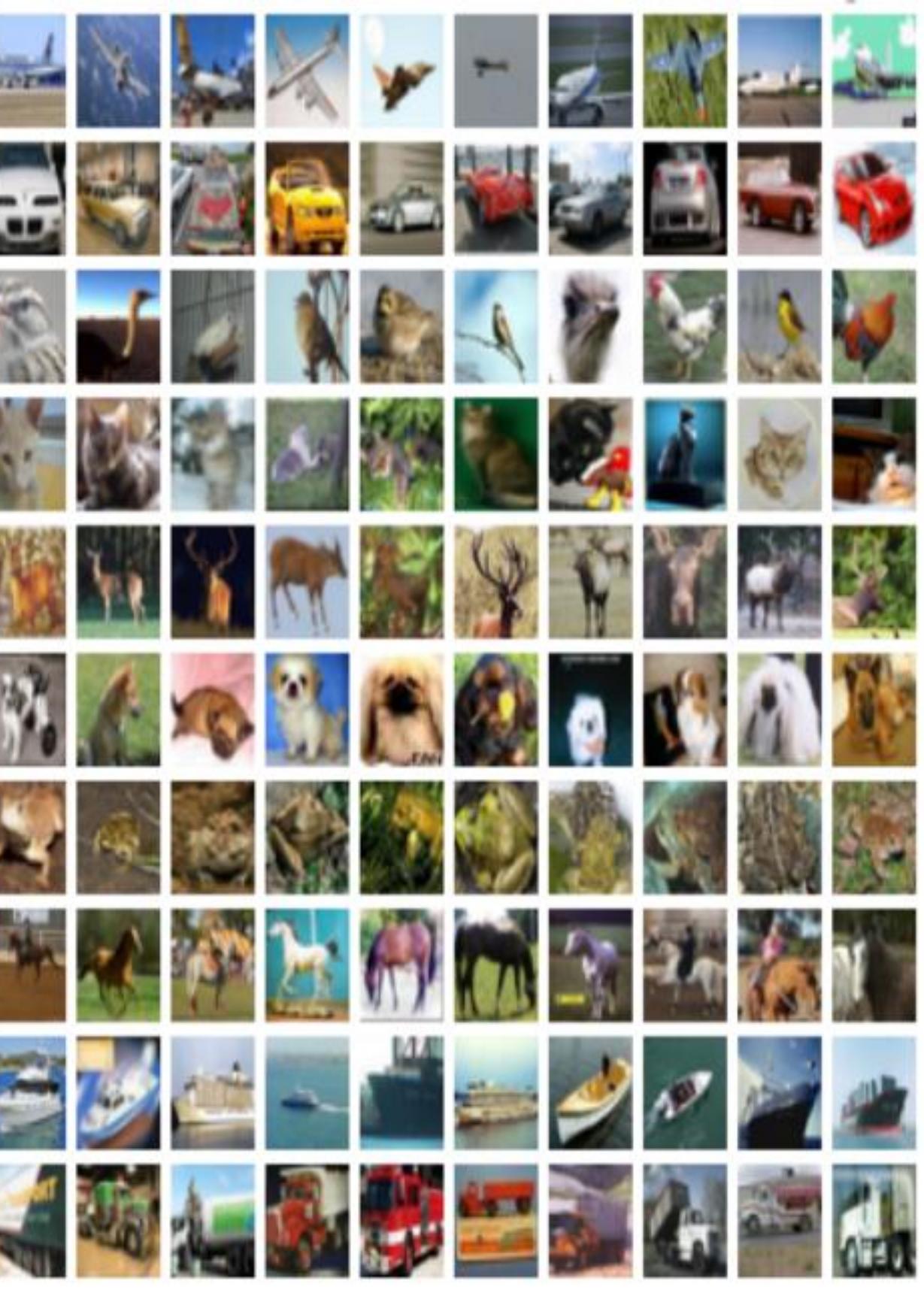
## Veri Seti Tanıtımı – CIFAR-10

- Sınıflar (10): airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck
- Her sınıf görsel açıdan birbirine yakın alt örüntüler taşıyabildiği için “inter-class confusion” analizi için uygundur.
- Bu çalışmada veri seti `torchvision.datasets.CIFAR10` ile otomatik indirilmiş, tek tip ön-işleme uygulanmıştır.

## Ön-İşleme (Ortak Protokol)

- Tüm modeller için giriş:  $224 \times 224$  (CIFAR-10  $32 \times 32 \rightarrow$  resize)
- Normalize: CIFAR-10 mean/std (kodda sabit)
- Amaç: tüm mimarileri aynı ölçekte karşılaştırmak

Not:  $32 \times 32 \rightarrow 224 \times 224$  büyütme, bilgi eklemez; sadece giriş ölçüğünü standardize eder.  
Bu nedenle performans; mimari kapasitesi + eğitim protokolü + augmentation ile belirlenir.



# Veri Bölümleme Stratejisi

CIFAR-10 veri setinin torchvision üzerinden sağlanan resmi train/test ayrimı kullanılmıştır.

Eğitim aşamasında data augmentation uygulanmış, performans metrikleri bağımsız test seti üzerinde raporlanmıştır.

## Eğitim Seti (Train)

50.000 görüntü (%83,33)

- Model ağırlıkları bu veri üzerinde öğrenilmiştir.
- Eğitim sırasında data augmentation (RandomResizedCrop, Horizontal Flip) uygulanmıştır.

## Data Augmentation (Yalnızca Train İçin)

50.000 görüntü (%83,33)

- Eğitim sırasında modelin genelleme kabiliyetini artırmak amacıyla
- aşağıdaki dönüşümler yalnızca train setine uygulanmıştır:
  - RandomResizedCrop (224×224)
  - RandomHorizontalFlip
  - Normalization (CIFAR-10 mean/std)
- Test verisi üzerinde herhangi bir augmentation uygulanmamıştır.

## Test Seti (Test)

10.000 görüntü (%16,67)

- Nihai başarı `test\_acc` bu veri üzerinde raporlanmıştır.
- Confusion matrix ve sınıf bazlı hata analizi bu aşamada yapılmıştır.

# Genel Yöntem Akışı

01

## Veri Yükleme ve Normalizasyon

CIFAR-10 indir → Tensor'a çevir → mean/std ile normalize

03

## Veri Artırma

Aug ON: RandomResizedCrop + HorizontalFlip

Aug OFF: Sadece Resize + Normalize

05

## Eğitim

AdamW optimizer, sabit lr (kodda lr=1e-3)

Varsayılan kısa eğitim: epochs=3 (rapor üretimi için)

07

## Test Değerlendirmesi

Test accuracy + confusion matrix

(Ayrıca classification\_report çıktısı terminalde üretilir)

02

## Giriş Boyutu Standardizasyonu

Tüm modeller 224×224 giriş görür (fair comparison)

04

## Model Seçimi

SimpleCNN, CustomCNN, ResNet-18, ViT-Tiny, ConvNeXt-Tiny

06

## Checkpoint Kaydı

En iyi val\_acc görülen model: best.pt

08

## Görselleştirme ve Karşılaştırma

Loss/accuracy eğrileri

Confusion matrix

Tüm modellerin tek grafikte karşılaştırması: compare\_all\_models.png

# ADİL KARŞILAŞTIRMA STRATEJİSİ

## Adil Karşilaştırma Protokolü: Aynı Şartlar

1

### Ortak Input Boyutu (224x224)

- Farklı mimarilerin farklı giriş bekłentisi olabildiği için tüm modeller tek boyuta sabitlendi.
- Amaç: performans farkını “giriş ölçüği” yerine “mimari özellikler” üzerinden okumak.

2

### Ortak Normalize ve Veri Yükleme

- CIFAR-10 mean/std ile normalize edilerek tüm modeller aynı dağılımda girdi görür.
- Bu, modelin eğitim sırasında ölçek farkına bağlı dengesizlik yaşamamasını azaltır.

3

### Ortak Eğitim Mantığı (Tek Pipeline)

- Aynı optimizer ailesi, benzer eğitim döngüsü ve aynı seed ile karşılaştırma yapılır (seed=42).
- Böylece deney koşulları daha kontrollü hale getirilir.

# VERİ ARTIRMA (AUGMENTATION) ve ABLATION (İki kolon)

## Veri Artırma ve Ön-İşleme Teknikleri



### Eğitim Seti Augmentation (Aug ON)

- RandomResizedCrop (scale=0.8–1.0): nesnenin farklı kırpım/ölçek varyasyonlarına karşı dayanıklılık
- HorizontalFlip: yatay simetri ile veri çeşitliliğini artırma
- Normalize: öğrenmeyi stabilize etme

### Aug OFF Senaryosu

- Sadece Resize(224×224) + Normalize
- Amaç: augmentation'ın katkısını net görmek (ablation)

### Bu Projede Ablation Nerede?

- ResNet-18 için Aug ON ve Aug OFF iki farklı koşul çalıştırıldı.
- Böylece “augmentation açıkken mi daha iyi genelliyor?” sorusuna deneysel yanıt üretildi.

# Değerlendirilen Model Mimarileri

Değerlendirilen Model Mimarileri (Toplam 5 Model)

Bu çalışmada CNN tabanlı klasik/özelleştirilmiş yapılar ile modern mimariler karşılaştırılmıştır.

## CNN Tabanlı Modeller

### SimpleCNN (Baseline)

- Az parametreli, hızlı öğrenen basit bir referans model
- AdaptiveAvgPool sayesinde giriş boyutuna bağımsız sınıflandırma başlığı

### CustomCNN (Geliştirilmiş CNN)

- BatchNorm + daha derin bloklar ile daha güçlü temsil öğrenme
- Baseline'a göre daha yüksek kapasite hedeflenir

### ResNet-18 (Residual CNN)

- Skip connection'lar ile daha stabil derin öğrenme
- Genelde güçlü bir genel amaçlı CNN omurgası olarak bilinir

## Transformer / Modern Mimari

### ViT-Tiny (Vision Transformer)

- Görüntüyü patch'lere bölüp attention ile temsil öğrenir
- Küçük boyutlu transformer varyantı ile "transformer yaklaşımı" test edilir

### ConvNeXt-Tiny (Modern CNN tasarımı)

- CNN dünyasında transformer-benzeri tasarım prensipleri
- Pratikte güçlü genelleme ve modern mimari avantajları hedeflenir

□ Adil Karşılaştırma Notu: Tüm modeller aynı giriş boyutu, aynı normalize ve aynı seed ile çalıştırılmıştır.

# Değerlendirme Metrikleri ve Önemi

Bu çalışmada model başarımı hem sayısal metriklerle hem de hata türlerini gösteren görsellerle değerlendirilmiştir.

## Accuracy (Doğruluk)

- Toplam tahminler içinde doğru sınıflandırılan örnek oranıdır.
- CIFAR-10 dengeli sınıf yapısına sahip olduğundan temel metrik olarak anlamlıdır.

## Test Accuracy

- Test: nihai raporlanan performans

## Confusion Matrix (Karışıklık Matrisi)

- Sınıf bazlı hangi sınıfların karıştığını gösterir.
- “Neden şu sınıflar karıştı?” yorumları bu görsel üzerinden yapılır.

## Classification report (Terminal)

- Precision/Recall/F1 gibi sınıf bazlı metrikleri verir (kodda eval sırasında üretilir).
- Raporlama için istenirse daha sonra dosyaya da kaydedilebilir.

Vurgu: Bu projede en kritik karşılaştırma çıktısı: tek protokol altında test accuracy sıralaması + confusion matrix yorumudur.

# METRICS SUMMARY TABLOSU

model	aug	acc	prec	rec	f1	auc	img	seed
resnet18	on	<b>0.7683</b>	0.7791	0.7683	0.7658	0.9747	224	42
resnet18	off	<b>0.7661</b>	0.7884	0.7661	0.7680	0.9753	224	42
convnext_tiny	on	<b>0.6282</b>	0.6436	0.6282	0.6200	0.9387	224	42
customcnn	on	0.5496	0.5936	0.5496	0.5422	0.9126	224	42
vit_tiny	on	0.4750	0.5060	0.4750	0.4602	0.8867	224	42
simplecnn	on	0.4353	0.4425	0.4353	0.4172	0.8625	224	42

## Detaylı Sıralama Yorumu

Bu deney koşullarında ResNet-18, hem augmentation açıkken hem kapalıyken en yüksek doğruluğu elde ederek açık ara lider görünmektedir. Bu durum, residual bağlantıların (skip connection) eğitim stabilitesini artırması ve küçük veri setlerinde güçlü genelleme sağlayabilmesi ile uyumludur. ConvNeXt-Tiny modern bir mimari olmasına rağmen kısa eğitim süresi (epochs=3) ve CIFAR-10'un 32×32'den 224×224'e büyütülmesi gibi etmenler altında ResNet kadar iyi yakınsamamış olabilir. ViT-Tiny'nin daha düşük doğruluğu ise transformer mimarilerinin genellikle daha uzun eğitim ve/veya daha büyük veri ihtiyacına sahip olabilmesiyle açıklanabilir.

# GENEL SONUÇ ÖZETİ

En İyi Modeller ve Sıralama

0.7683

ResNet-18 (Aug ON)

Test Accuracy

0.7661

ResNet-18 (Aug OFF)

Test Accuracy (Ablation)

0.6282

ConvNeXt-Tiny (Aug ON)

Test Accuracy

## Detaylı Metrik Karşılaştırması

Model	Acc	Kısa Yorum
ResNet-18 Aug ON	0.7683	En iyi genel performans
ResNet-18 Aug OFF	0.7661	Aug katkısı bu koşulda sınırlı
ConvNeXt-Tiny	0.6282	Modern CNN, daha fazla eğitimle artabilir
CustomCNN	0.5496	Baseline üstü, ama sınırlı genelleme
ViT-Tiny	0.4750	Daha fazla epoch/veri ile iyileşebilir
SimpleCNN	0.4353	Baseline; kapasite sınırlı

# ABLATION: RESNET-18 AUG ON vs OFF

Ablation Analizi: Data Augmentation (ResNet-18)



aResNet-18 (Aug OFF)



aResNet-18 (Aug ON)

## Yorum:

- Ablation sonucu, bu deney koşullarında augmentation'in ResNet-18 için doğruluğu yalnızca marjinal düzeyde artırdığını göstermektedir.
- Bunun birkaç olası açıklaması vardır:
  1. eğitim süresi çok kısa olduğu için (epochs=3) augmentation'ın genelleme katkısı tam ortaya çıkmamış olabilir;
  2. CIFAR-10'un zaten dengeli ve sınıf sayısı düşük bir veri seti olması augmentation ihtiyacını sınırlamış olabilir;
  3. kullanılan augmentation seti (RandomResizedCrop + HorizontalFlip) nispeten hafif olduğundan etkisi küçük kalmış olabilir.
- Buna rağmen Aug ON senaryosu daha iyi olduğu için, daha uzun eğitimde farkın büyümeye olasılığı vardır.

# CNN vs Transformer/Modern: Karşılaştırmalı Analiz

## En İyi Performans: CNN Taraflı

### ResNet-18 (0.7683 test acc)

- Residual bağlantılar sayesinde hızlı ve stabil yakınsama
- Kısa eğitimde bile güçlü performans
- CIFAR-10 gibi klasik benchmark'ta güvenilir bir referans

## Transformer/Modern Taraflı

### ConvNeXt-Tiny (0.6282) ve ViT-Tiny (0.4750)

- Bu koşullarda ResNet'in gerisinde kalmıştır.
- Olası neden: Transformer'lar genelde daha uzun eğitim ve daha büyük veriyle daha iyi açılır.
- ConvNeXt, modern tasarıma rağmen kısa eğitimde ResNet kadar iyi optimize olamamış olabilir.a

**Çıkarım:** Bu deneyde “en iyi mimari” sonucu, yalnız mimari değil; eğitim süresi, veri boyutu, augmentation ve optimizasyon dinamiklerinin birlikte etkisidir.a

# CONFUSION MATRIX OKUMASI

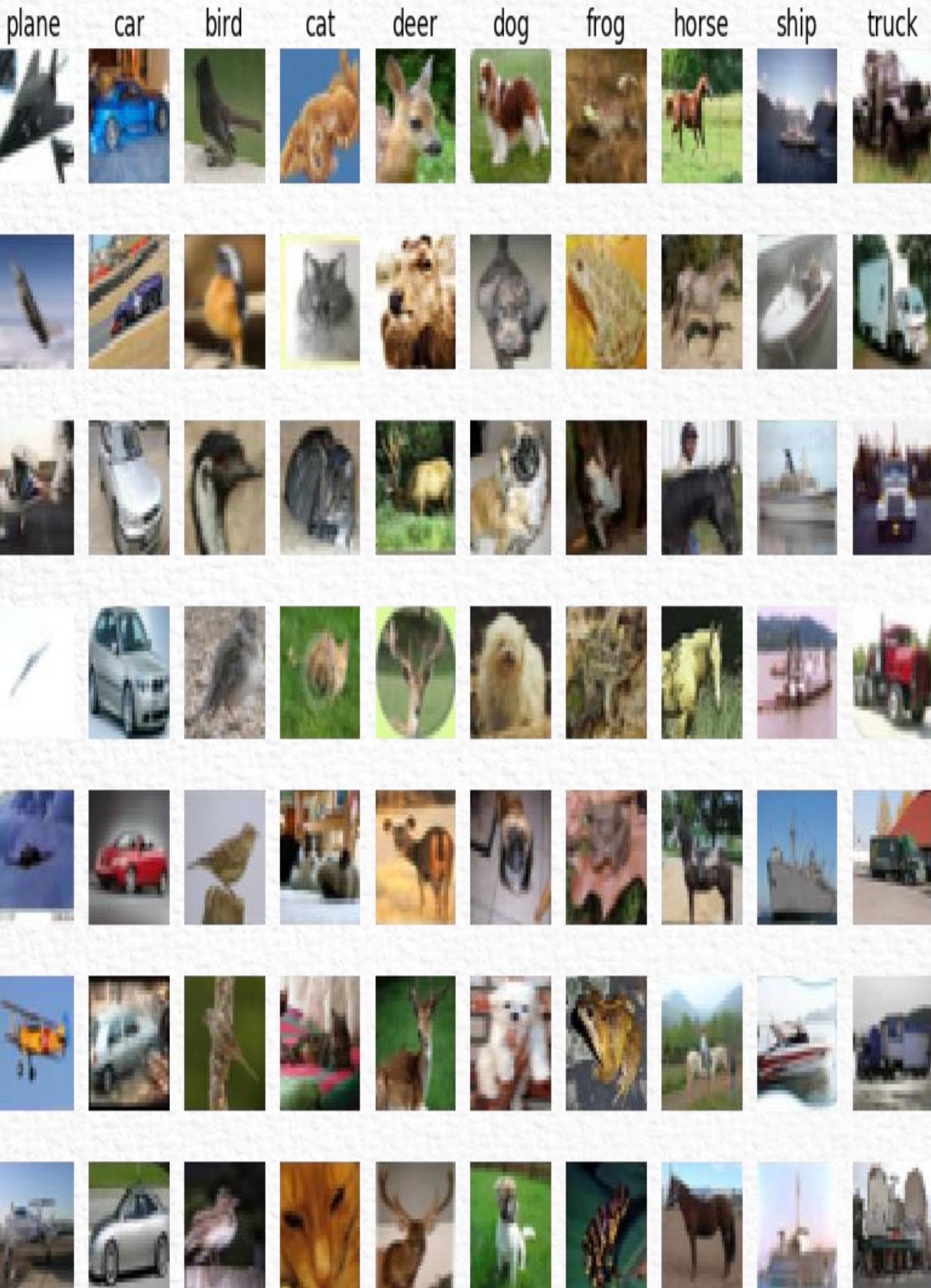
## Confusion Matrix Üzerinden Hata Analizi (Sınıf Bazlı)

- Bu slaytta her modelin confusion matrix görseli üzerinden sınıf karışmalarının nedeni yorumlanır. CIFAR-10 sınıflarında bazı çiftler görsel olarak birbirine yakındır ve benzer arka plan/poz varyasyonları karışmayı artırır.

## Tipik karışan sınıflar (sunum diliyle uzun yorum)

- acat ↔ dog: Doku ve poz benzerliği, küçük çözünürlükte ayırt edici detayların kaybolması
- deer ↔ horse: Benzer siluet ve doğal arka planlar
- automobile ↔ truck: Taşıt sınıflarında perspektif ve ölçek farkı
- ship ↔ airplane: Bazı görüntülerde arka plan (gökyüzü/deniz) ve düşük çözünürlük etkisi

- Kritik Nokta:** Giriş boyutu  $224 \times 224$  olsa bile bilgi kaynağı  $32 \times 32$ 'den geldiği için bazı ince ayrımlar geri getirilemez. Bu nedenle confusion matrix, modelin hangi görsel ipuçlarını yakalayabildiğini dolaylı olarak gösterir.



# Tartışma: Sonuçlar Neden Bu Şekilde?

## Açıklamalar

### 1 Kısa Eğitim (epochs=3) Etkisi

- 3 epoch, özellikle transformer/modern mimariler için yetersiz olabilir.
- ResNet gibi “optimizasyonu kolay” mimariler kısa eğitimde avantajlıdır.

### 2 CIFAR-10'un Düşük Çözünürlüğü ve Resize

- $32 \times 32 \rightarrow 224 \times 224$  standardizasyon adil karşılaştırma sağlar, ancak bilgi eklemez.
- Bazı sınıflar arasındaki ince farklar orijinal çözünürlükte zaten sınırlıdır.

### 3 Model Kapasitesi ve İndüktif Önyargı (Inductive Bias)

- CNN'ler lokal örüntülerini (kenar/doku) yakalamada güçlü doğal önyargıya sahiptir.
- ViT gibi attention tabanlı yapılar, yeterli veri/eğitim olmadan bu avantajı kısa sürede yakalayamayabilir.

Bu çalışma, “nihai en iyi skor” yarışından çok, aynı protokol altında mimari farkını görür kılan karşılaştırmalı analiz üretmeyi hedefler.

# Sonuçlar ve Ana Çıkarımlar

## 1 En İyi Model: ResNet-18

- Test accuracy: 0.7683 (Aug ON)
- Kısa eğitimde bile en iyi yakınsama ve genelleme

## 2 Augmentation Ablation (ResNet-18)

- Aug ON vs OFF farkı küçük (0.7683 vs 0.7661)
- Daha uzun eğitimde/çeşitli augmentation ile fark büyüyebilir

## 3 Modern Mimari: ConvNeXt-Tiny Orta Seviye

- Test accuracy: 0.6282
- Potansiyel olarak daha fazla epoch, lr schedule ve/veya pretrained ile artış beklenebilir

## 4 ViT-Tiny ve Baseline'lar

- ViT-Tiny: 0.4750, SimpleCNN: 0.4353
- Bu koşullarda transformer yaklaşımı belirgin avantaj göstermemiştir

## Model Karşılaştırması Özeti

1

### En güvenilir seçim

- Kısa eğitim + CIFAR-10'da ResNet-18 en güçlü profil

2

### Aug etkisi

- Bu protokolde sınırlı; daha uzun eğitimde yeniden test edilmeli

3

### Modern mimariler

- ConvNeXt umut verici, ViT daha fazla veri/epoch ister

# Gelecek Çalışmalar ve Öneriler

1

## Epoch artırımı ve LR schedule

- 3 epoch yerine 20–50 epoch aralığında eğitim
- Cosine annealing / step decay gibi schedule ile daha iyi yakınsama

2

## Çoklu seed ile stabilité analizi

- seed=42 tek koşul yerine 3–5 farklı seed ile ortalama  $\pm$  std raporu
- “model gerçekten daha iyi mi, yoksa şans mı?” sorusuna daha güçlü yanıt

3

## Pretrained ağırlıklarıyla transfer learningx

- Özellikle ViT/ConvNeXt gibi mimariler için pretrained başlatma
- Küçük veri setinde genelleme artışı beklenir

4

## Daha zengin augmentation ve regularizationa

- RandAugment, CutMix/MixUp, RandomErasing gibi tekniklerin etkisi
- Özellikle overfitting kontrolü ve robustluk için

# Tesekkürler