

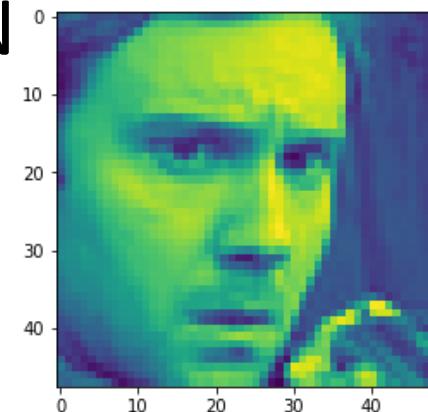
Ham Piksel Tabanlı Yüz İfadesi Tanıma CNN ile 8 Sınıflı Duygu Sınıflandırması

**Student: YASİR ÖMER ALPARSLAN
Student ID: 244225708**

Danışman: Öğretim Üyesi
Dr. SELİM YILMAZ
MUĞLA ÜNİVERSİTESİ / FEN BİLİMLERİ ENSTİTÜSÜ / Bölüm:YZ
Tarih: 21-12-2025

Problem Tanımı ve Motivasyon

- Yüz ifadesi tanımanın İnsan–bilgisayar etkileşimi HCI, güvenlik, sağlık ve eğitim gibi alanlarda yüz ifadesi tanımanın önemi
- Klasik yöntemler: HOG, LBP, SIFT + SVM (manuel öznitelik çıkarımı)
- Bu çalışmada yalnızca ham pikseller ve Tamamen derin öğrenme tabanlı CNN kullanılmıştır



Çalışmanın Amacı

- 48×48 gri seviye yüz görüntülerinden duyu tanıma
- 8 sınıf **8 duyu sınıfını** tahmin etmek:

Anger, Contempt, Disgust, Fear, Happiness, Neutral, Sadness, Surprise
(Öfke, Küçümseme, Tiksinti. Korku. Mutluluk. Tarafsızlık. Üzüntü, Şaşkınlık)



- Uçtan uca (end-to-end), hafif ve tekrar üretilenbilir CNN modeli.
- **Hiçbir manuel özellik çıkarımı yapılmamıştır.**
- Model yalnızca **ham pikseller** üzerinden öğrenmiştir.
- **Matematiksel olarak:**
 - Her görüntü şu şekilde temsil edilir:
 $\mathbb{R}^{1 \times 48 \times 48} \ni X$
 - Daha sonra tüm pikseller normalize edilmiştir:
 $\frac{X}{255} = {}_{norm}X$
 - Bu adım, gradyanların daha kararlı olmasını sağlar.

Modelin:

- Yeterli doğrulukta (%55 ≈ test doğruluğu)
- Hafif ve tekrar üretilenbilir olması (Keras kodu + kayıtlı ağırlıklar)

Veri Seti

- 48×48 boyutunda gri seviye yüz görüntüleri
- Klasör yapısı: train/ (0–7) 8 , test/ (0–7) 8
- Test seti: toplam 3,589 görüntü
- The training set consists of 28,709 examples
- <https://www.kaggle.com/code/pedroadorigel/lo/gradcam-fer2013-test/input>



Anger



Disgust



Fear



Happiness



Sadness



Surprise

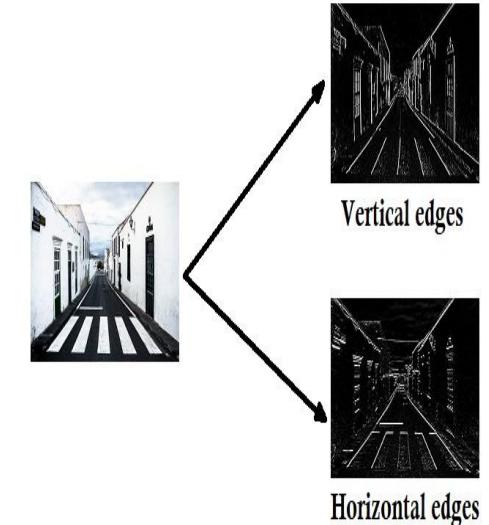


Neutral



Ön İşleme ve Veri Artırma

- Normalizasyon: $0-255 \rightarrow 0-1$
- One-hot label encoding (8 sınıf)
- %80 eğitim – %20 doğrulama
- Data augmentation(Veri büyütme): döndürme, kaydırma, horizontal flip
- Çok küçük ve gri tonlamalı görüntüler, bu da görevi zorlaştırır ancak modelin boyutunu küçültür.



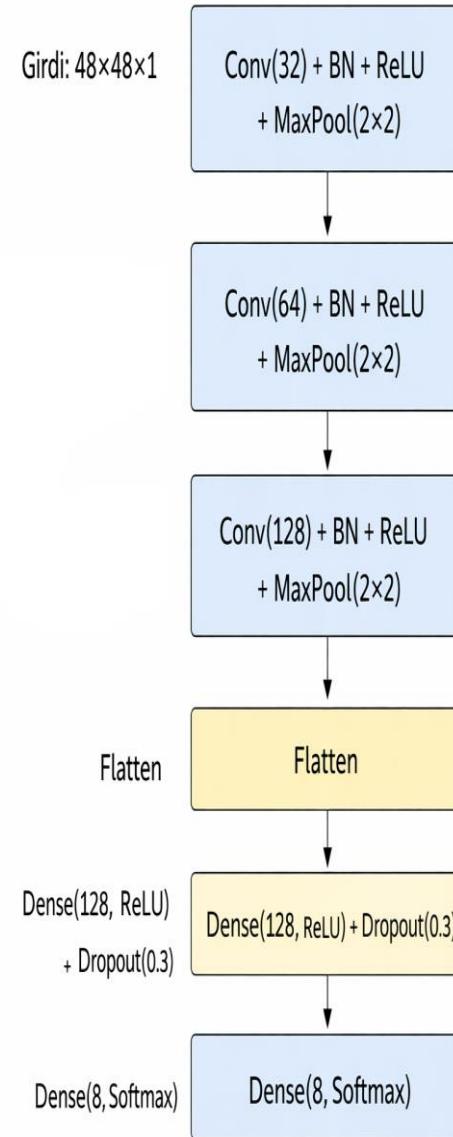
Önce

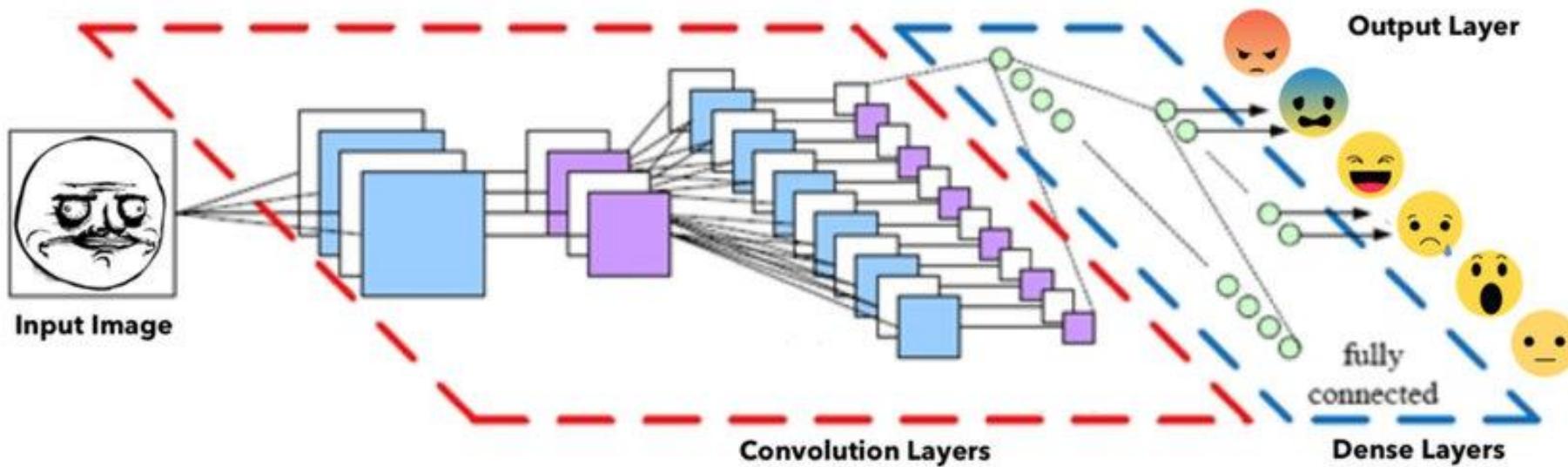
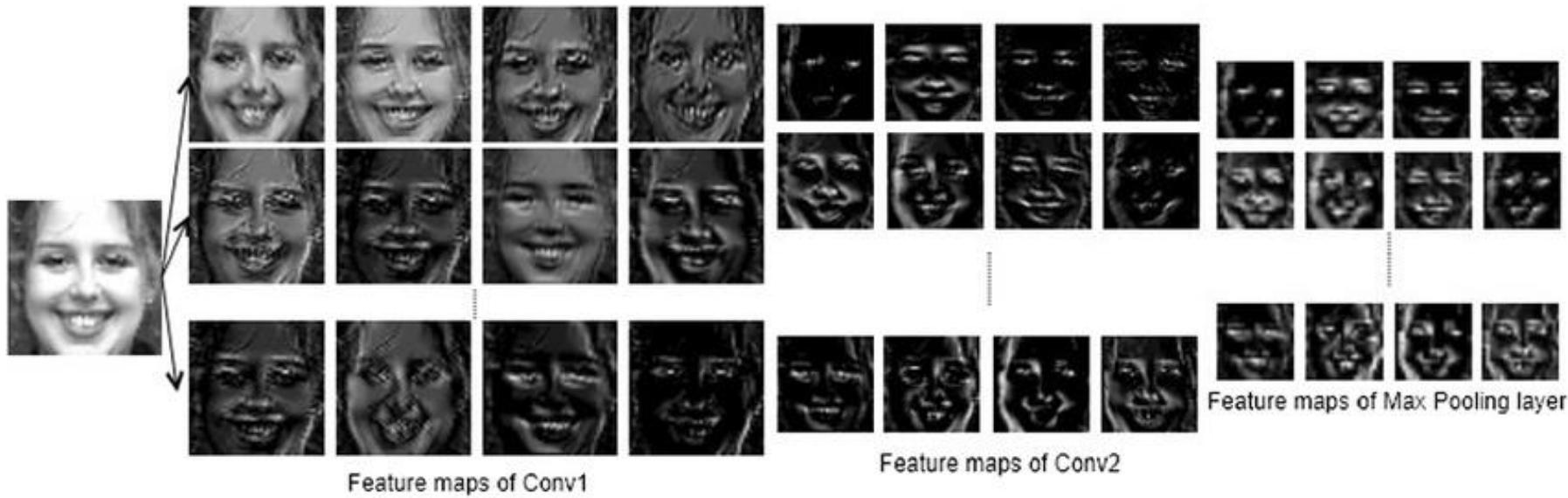


sonra

CNN Mimarisi

- Girdi: $48 \times 48 \times 1$
- Conv(32) Dönüşüm → Conv(64) Dönüşüm → Conv(128) Dönüşüm + MaxPool
- Dense(128) MaksHavuzYoğun + Dropout(0.3) Eksik
- Çıkış: Dense* Yoğun*(8, Softmax)





Eğitim Kurulumu

- Batch size: 64, Epoch: 15,60,360. Optimizer: Adam ($lr = 1e-3$)
(EarlyStopping yok, tüm epoch'lar tamamlandı)
- ModelCheckpoint **best_emotion_cnn_8class.keras**
- ve Reduce LROn Plateau **val_loss** göre öğrenme oranını düşürdü bunları ikisini kullanıldı.
- **ResNet veya transfer** öğrenme yöntemlerini kullanmadı; bunun yerine, sıfırdan oluşturulmuş hafif bir model kullanıldı.

Bu adımlar, modelin sadece tren sayısını azaltmakla kalmayıp genelleme yapmasına da yardımcı olur.

CNN Modeli Ham Görüntülerden Ne Öğrendi?

CNN modeli, yüzü **anlamlandırmaz**, onu **hiyerarşik görsel örüntülere** ayırarak öğrenir.

Öğrenilen Bilgi	Katman
Kenarlar, kontrast	İlk Conv
Göz, ağız, kaş bölgeleri	Orta Conv
Duygusal ifade örüntüleri	Son Conv

Konvolusyon işlemi (matematiksel):

$$W(m, n) \cdot (n + m, j + X(i \sum_{m,n} = X)(i, j) * F)$$

Bu işlemden sonra:

Batch Normalization → dağılım stabilitesi

ReLU:

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

MaxPooling → boyut indirgeme + gürültü azaltma

Karar Mekanizması: Softmax

Son tam bağlantılı katmanda her sınıf için bir skor üretilir:

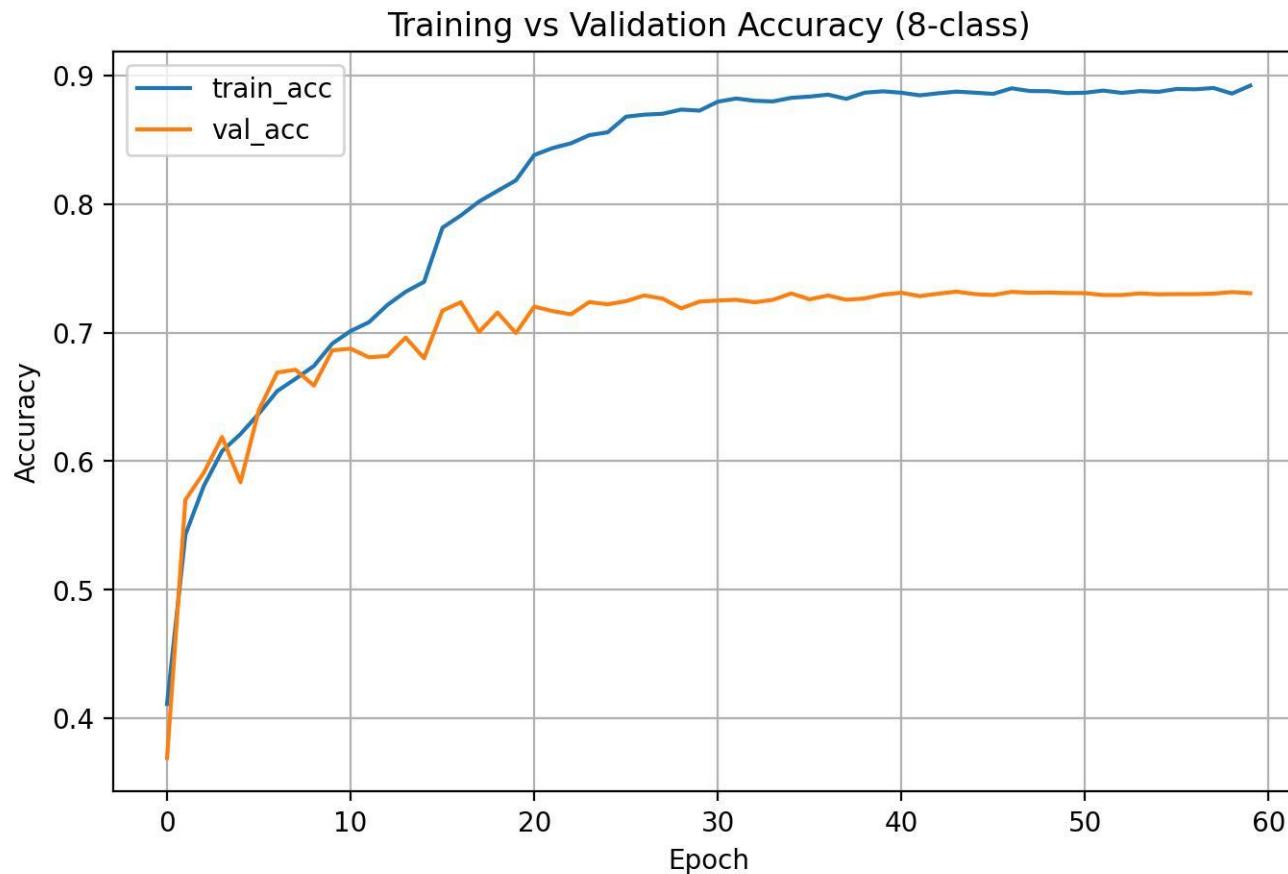
$$b + Wx = z$$

Ardından Softmax uygulanır

$$\frac{e^z}{\sum_{j=1}^k e^z} = P(y=k|x)$$

Model, en yüksek olasılığa sahip sınıfı tahmin olarak seçer.

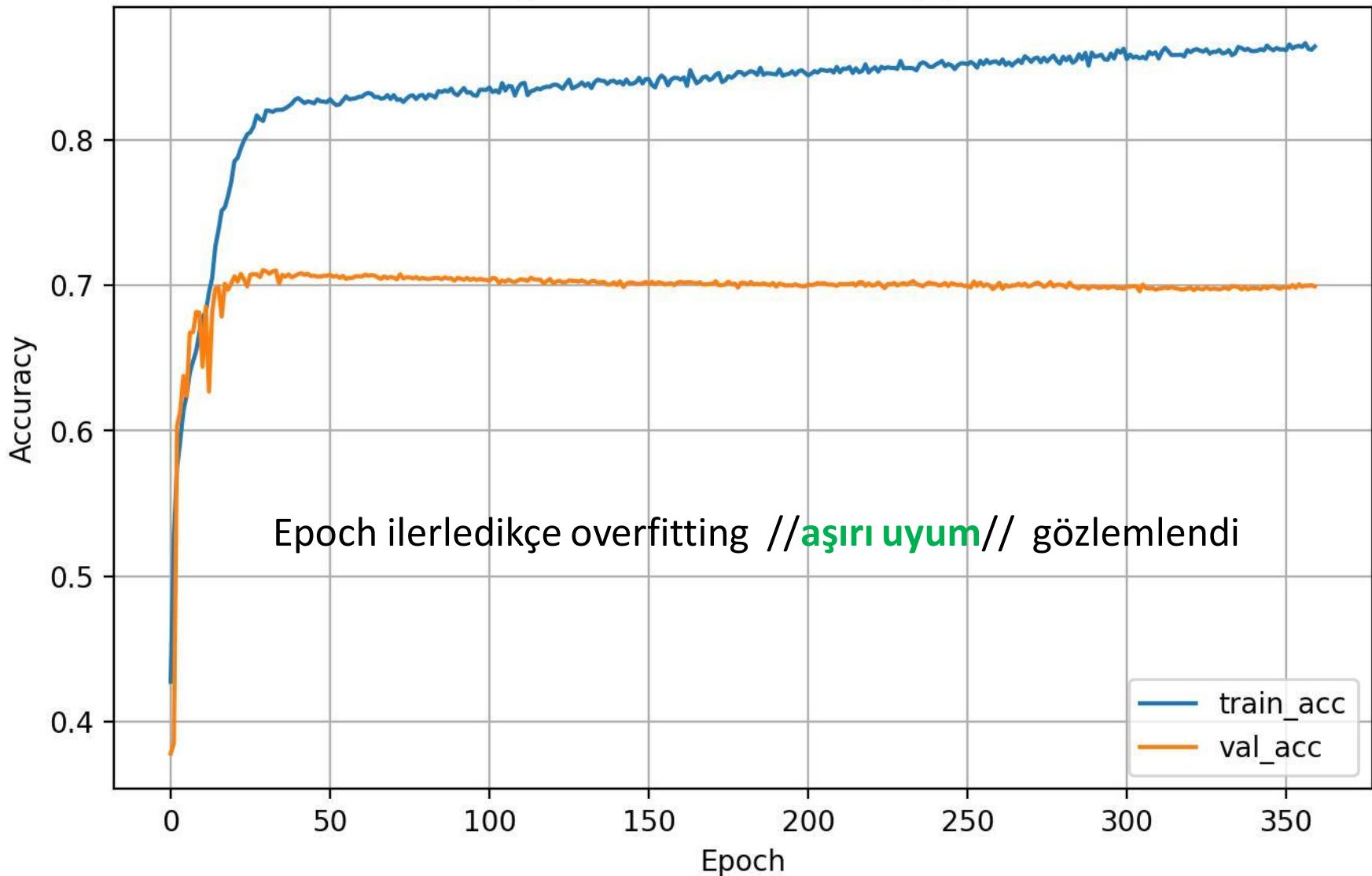
Eğitim ve Doğrulama Doğruluğu



- Training accuracy: %43 → %90+
- Validation accuracy //Doğrulama doğruluğu//: %40 → %75
- Epoch ilerledikçe overfitting// aşırı uyum // gözlemlendi

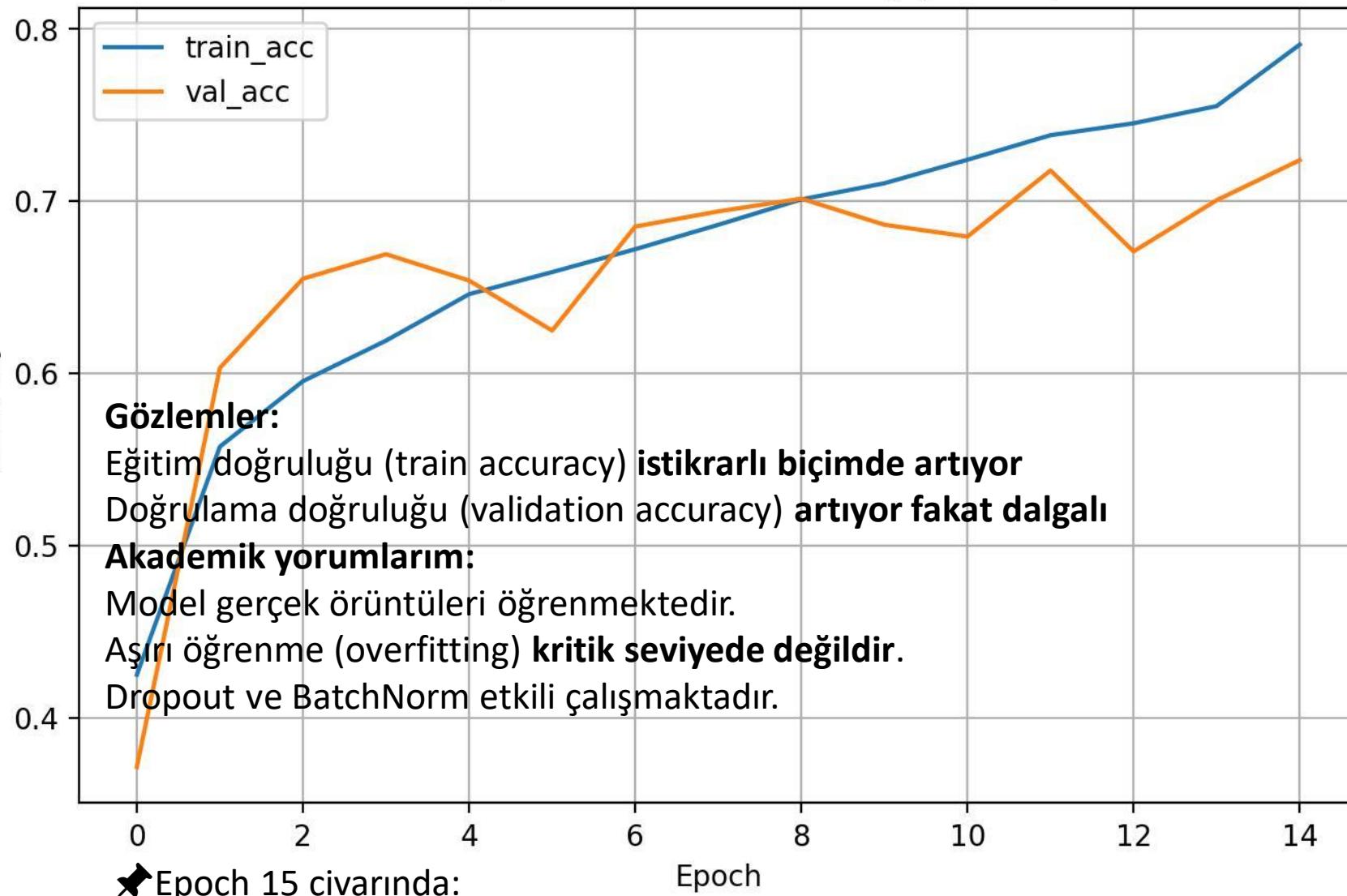
Eğitim ve Doğrulama Doğruluğu

Training vs Validation Accuracy (8-class)



Eğitim ve Doğrulama Doğruluğu

Training vs Validation Accuracy (8-class)



Gözlemler:

Eğitim doğruluğu (train accuracy) **istikrarlı biçimde artıyor**

Doğrulama doğruluğu (validation accuracy) **artıyor fakat dalgalı**

Akademik yorumlarım:

Model gerçekörüntülerini öğrenmektedir.

Aşırı öğrenme (overfitting) **kritik seviyede değildir.**

Dropout ve BatchNorm etkili çalışmaktadır.

★ Epoch 15 civarında:

Epoch

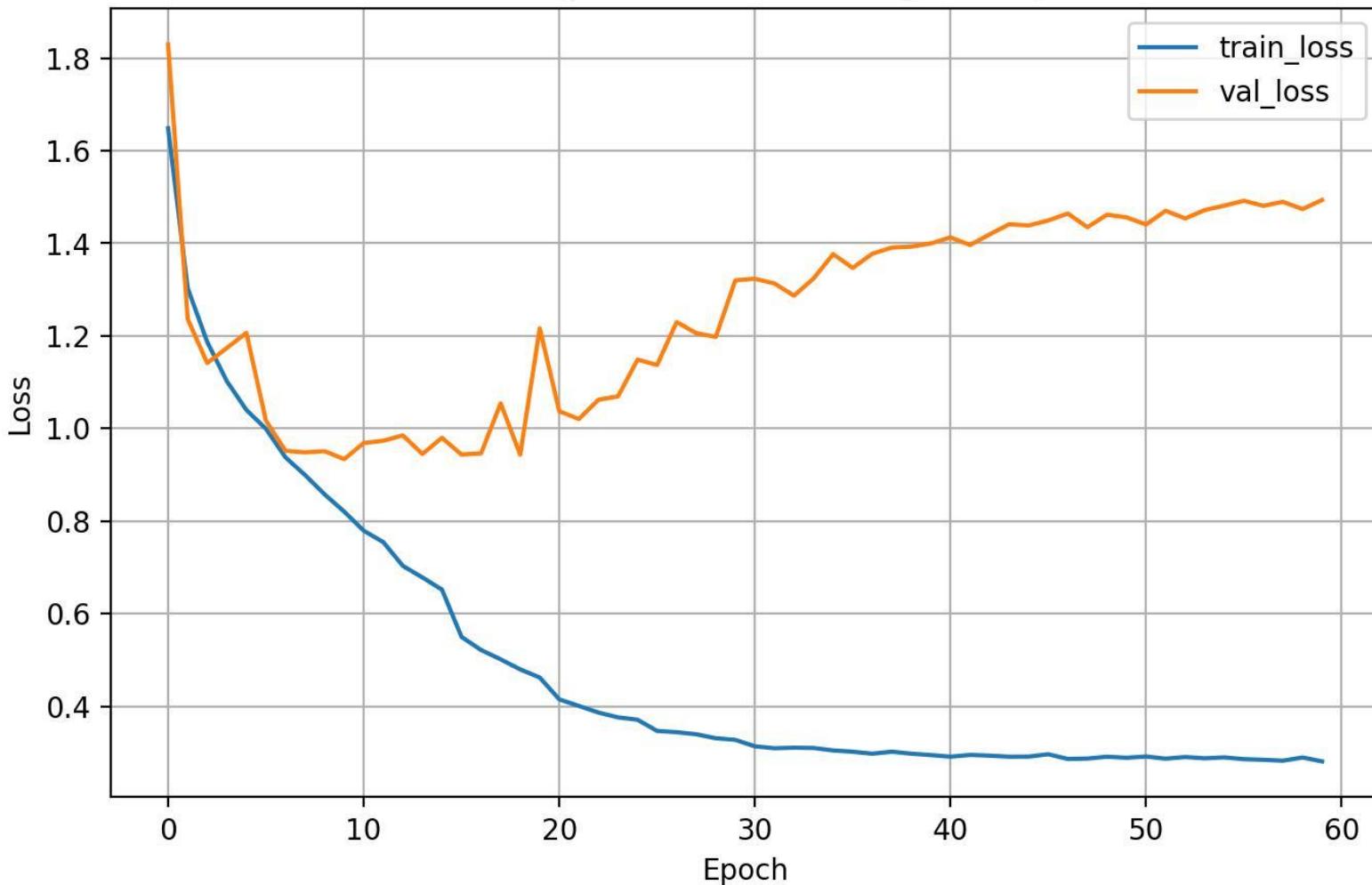
Train accuracy \approx %79

Validation accuracy \approx %72

Bu, modelin **öğrenmeye devam edebilecek potansiyele sahip** olduğunu göster

Eğitim ve Doğrulama Kaybı

Training vs Validation Loss (8-class)

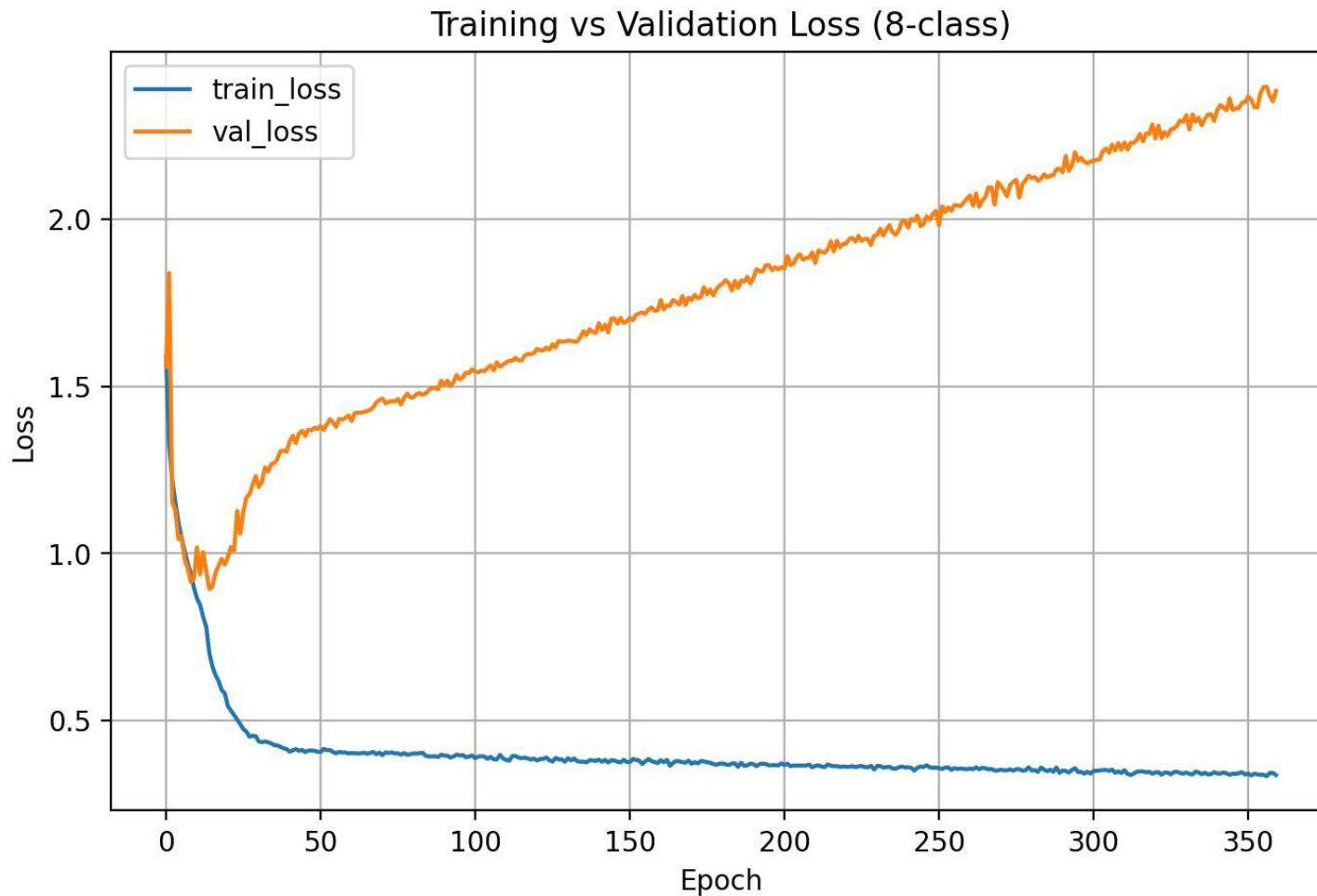


- Training loss: $\sim 1.65 \rightarrow 0.16$
- Validation loss: en iyi nokta epoch 10–15
- Sonrasında validation loss artışı (overfitting //**aşırı uyum**//)

Kullanılan kayıp fonksiyonu:

$$(\hat{y}_k y_k \log \sum_{k=1}^8 -) = \mathcal{L}$$

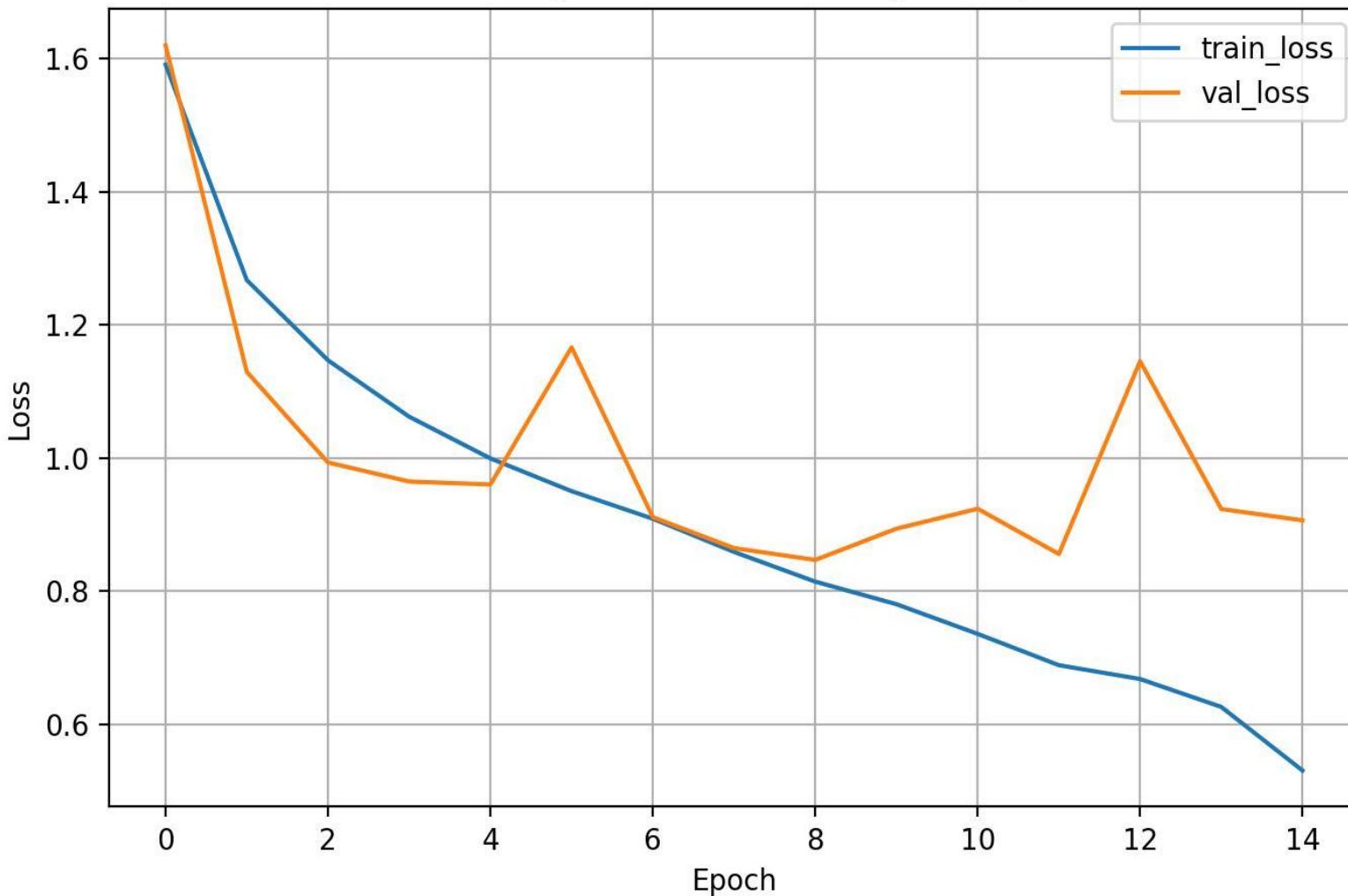
Eğitim ve Doğrulama Kaybı



- **Gözlemler:**
- Eğitim kaybı düzenli azalıyor ve Doğrulama kaybı genel olarak azalıyor ancak dalgalanıyor
- **Anlamı:** Model bazı sınıflarda kararsızlık yaşamaktadır. Ayrıca Bu durum **veri dengesizliği ve sınıf belirsizliği** ile ilişkilidir.

Eğitim ve Doğrulama Kaybı

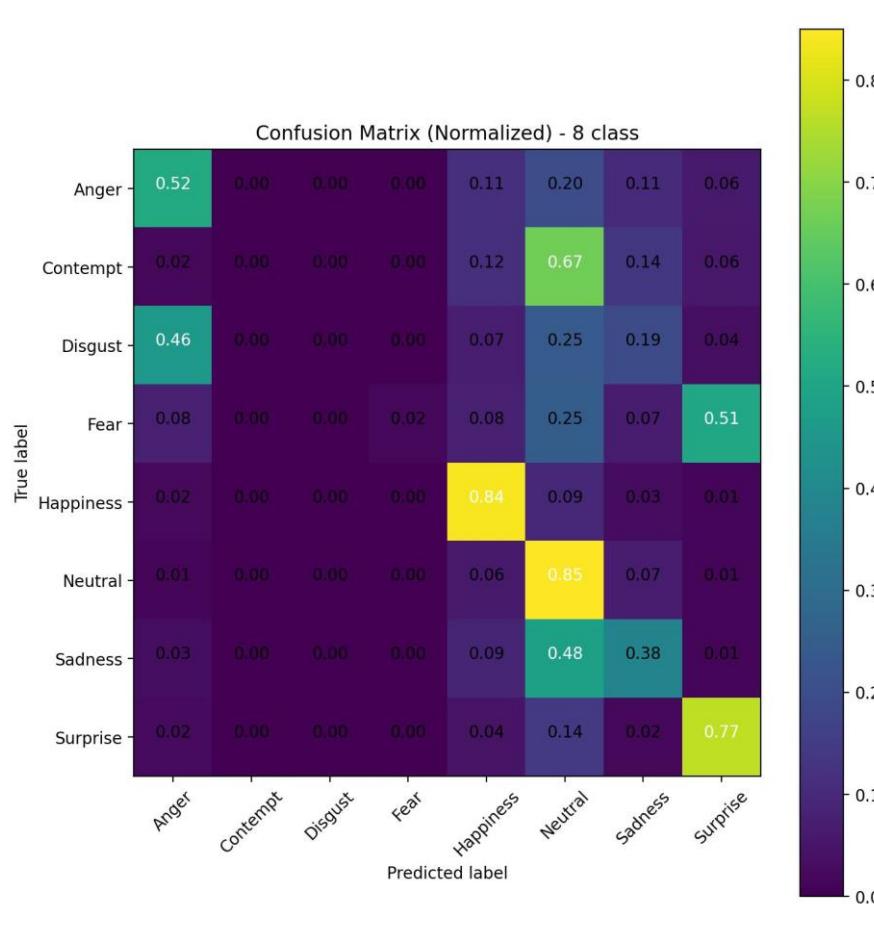
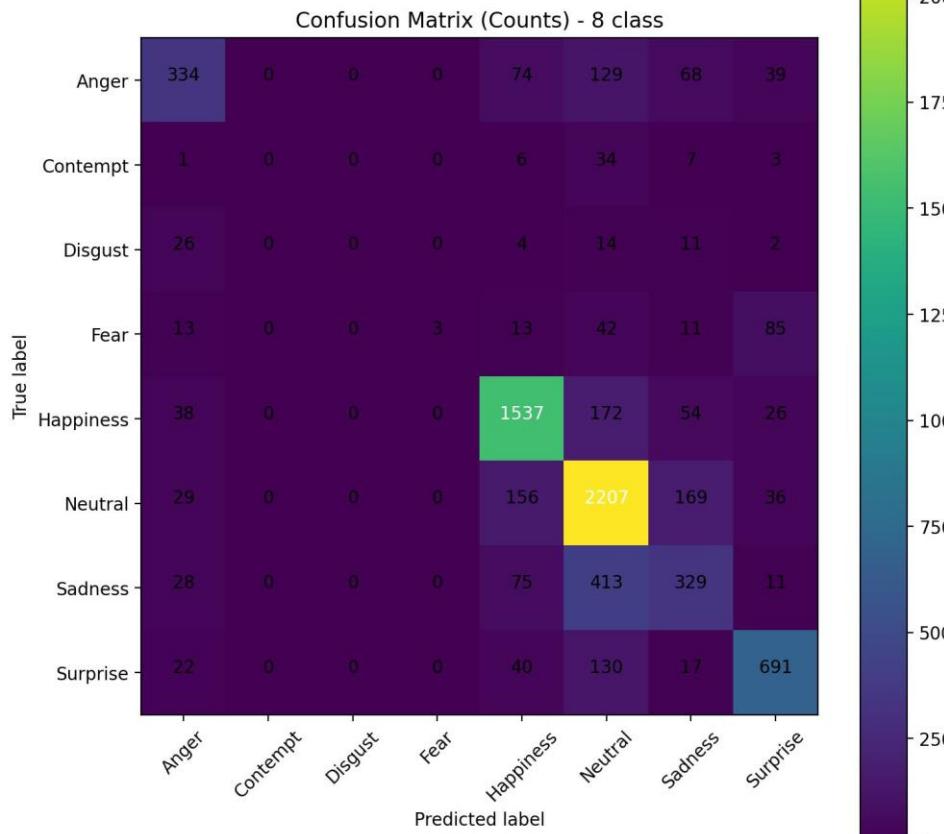
Training vs Validation Loss (8-class)



- Training loss: $\sim 1.6 \rightarrow 0.16$
- Validation loss: en iyi nokta epoch 10–15

Karmaşıklık Matrisi Confusion Matrix (Sayısal) – Nerede Hata Yapıldı?

- Test doğruluğu: %71.6 for 60 epoch , Test Loss: 1.5164 Test Accuracy: 0.7186
- En iyi sınıflar: Happiness, Neutral, Surprise, Anger
- Zayıf sınıflar: Contempt, Disgust, Fear



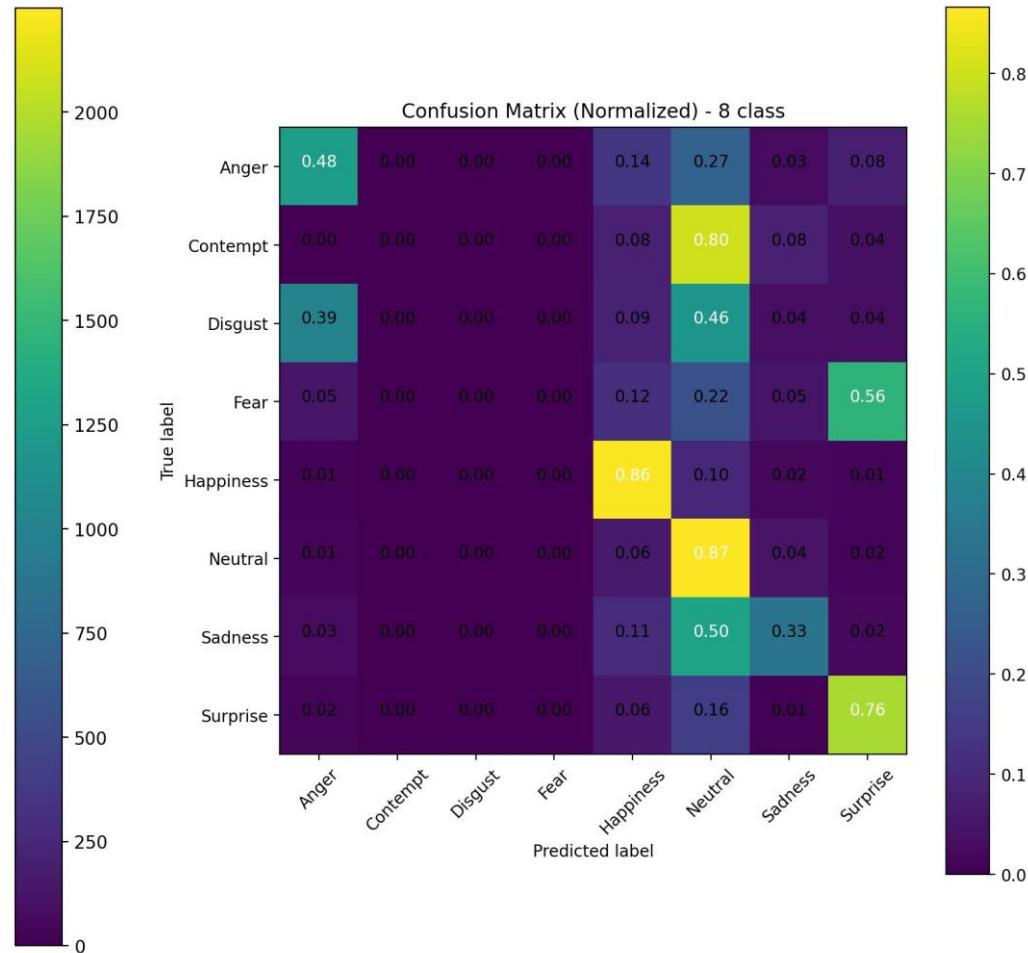
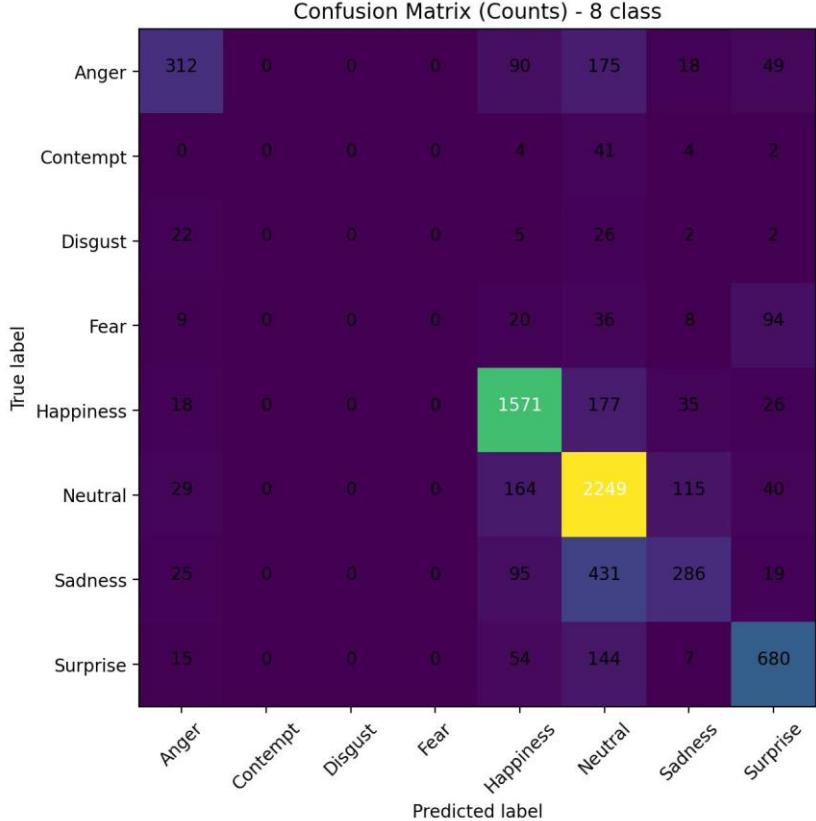
Model özellikle: Anger \leftrightarrow Neutral , Fear \leftrightarrow Surprise
arasında karışıklık yaşamaktadır.

Karmaşıklık Matrisi

- Test doğruluğu: %71.8 Test Loss: 0.9193 Test Accuracy: 0.7181 for 10-15 epoch

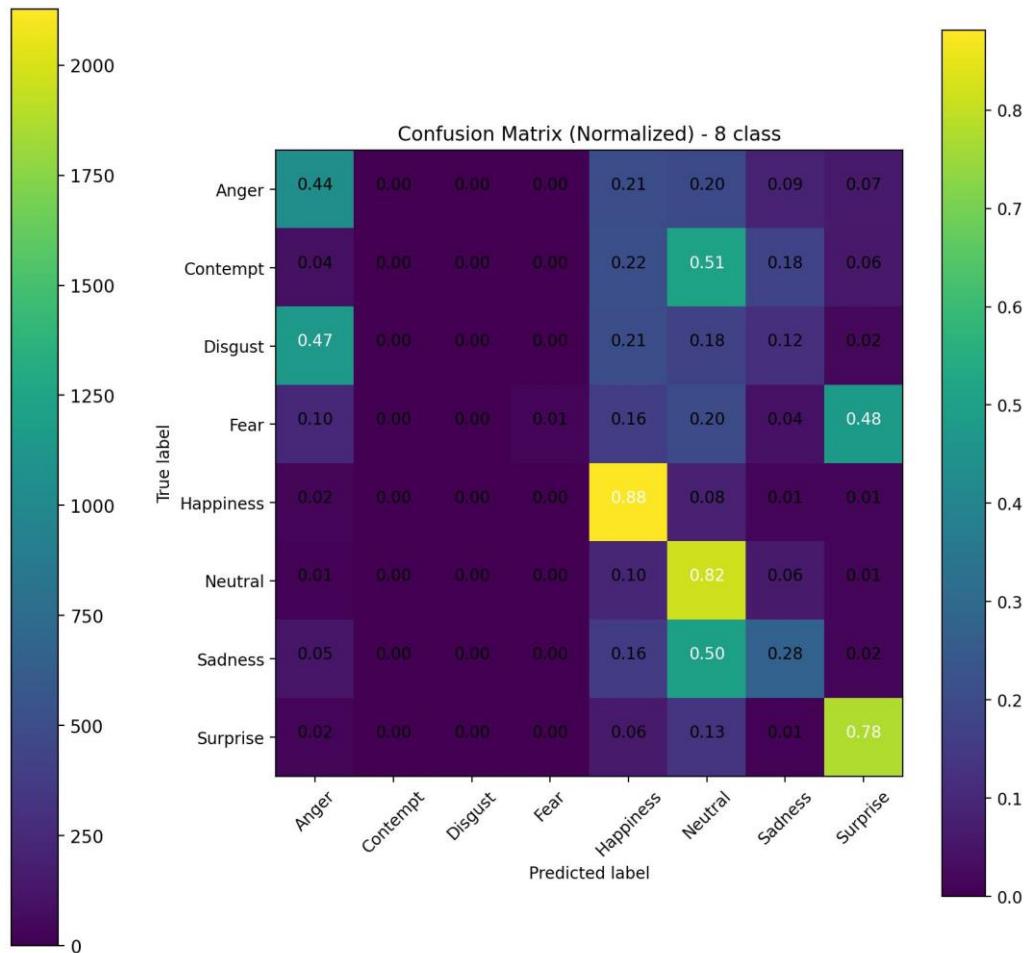
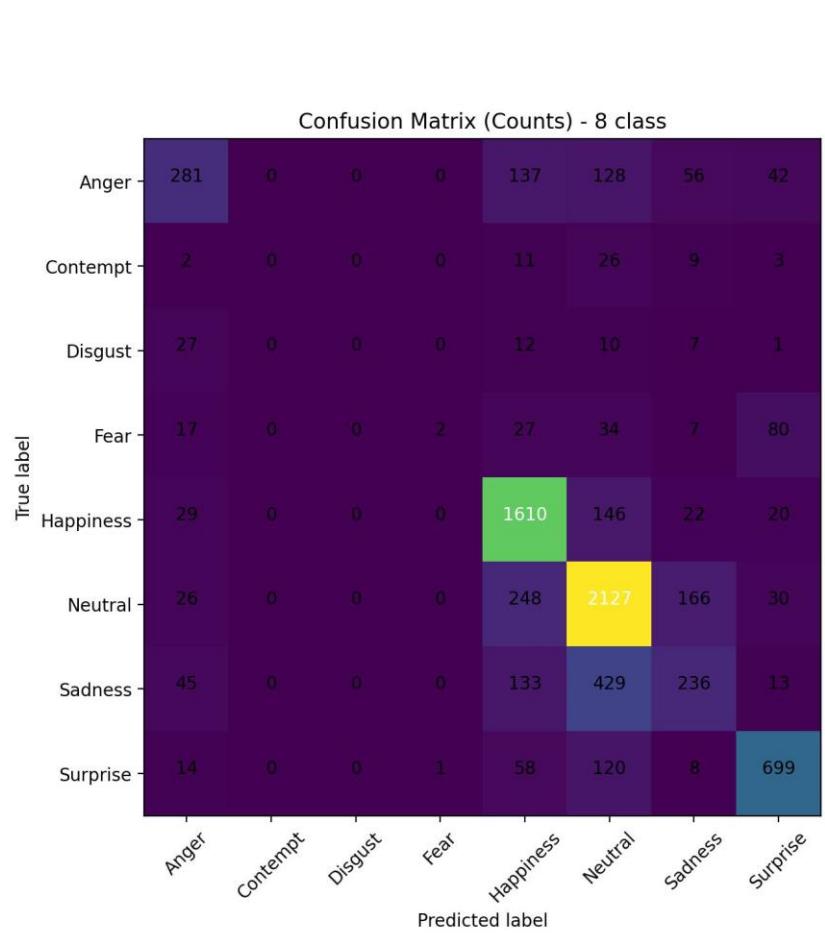
Yorumu:

- Happiness ve Neutral **belirgin görsel örüntülere sahiptir**
- Fear ve Disgust gibi ifadeler:
 - Görsel olarak belirsiz
 - Düşük çözünürlükte ayırt edilmesi zor
- Bu durum **model hatası değil**, problem doğasının bir sonucudur.



Karmaşıklık Matrisi

- Test doğruluğu: %54.5 Test Loss: 2.4276 Test Accuracy: 0.6980 for 360 epoch
- En iyi sınıflar: Happiness, Neutral, Surprise, Anger
- Zayıf sınıflar: Contempt, Disgust



Grad-CAM Nedir? (Teorik Açıklama)

Grad-CAM, modelin **hangi bölgelerden etkileneerek karar verdiği**ni bilmeni sağlar.
Amaç:

Model gerçekten yüzün anlamlı bölgelerine mi bakıyor!

Grad-CAM Matematiksel Tanımı

Son konvolüsyon katmanının özellik haritaları:

$$^{u \times v} \mathbb{R} \ni {}^k A$$

Hedef sınıf skoru:

$${}^c y$$

Ağırlıklar:

$$\frac{{}^c y \partial}{{}^k A \partial} \sum_j \sum_i \frac{1}{Z} = {}^c k \alpha$$

Grad-CAM haritası:

ReLU, yalnızca **pozitif katkıları** tutar.

$$\text{ReLU} \left(\sum_k \alpha_k^c A^k \right) = {}^c_{\text{Grad-CAM}} L$$

Örnek Tahmin Görselleri – Neden Yanıldı?

CNN:

- Ağız açıklığını → Happiness
- Yüz stabilitesini → Neutral
- olarak yorumlayabilmektedir.

✗ CNN:

- Psikolojik bağlamı **anlamaz**
- Sadece piksel örüntülerine bakar



Örnek Tahminler

- Doğru ve yanlış sınıflandırma örnekleri
- Görsel benzerlikten kaynaklanan hatalar
- Anger – Neutral – Happiness karışmaları

T:Anger
P:Happiness



T:Anger
P:Anger



T:Anger
P:Anger



T:Anger
P:Neutral



T:Anger
P:Sadness



T:Anger
P:Happiness



T:Anger
P:Anger



T:Anger
P:Neutral



T:Anger
P:Anger



T:Anger
P:Neutral



T:Anger
P:Anger



T:Anger
P:Neutral



T:Anger
P:Sadness



T:Anger
P:Sadness



T:Anger
P:Anger



T:Anger
P:Neutral



Sonuçlar

- Ham piksel tabanlı CNN ile 8 sınıfı duyu tanıma
- Test doğruluğu \approx %54.5 tüm ham veri epoch 15 için Basit ama etkili bir mimari
- Bu çalışmada geliştirilen CNN tabanlı duyu sınıflandırma modeli, **Mutluluk**, **Şaşkınlık** ve **Tarafsızlık** gibi duygularda görece yüksek başarı göstermesine rağmen, **Korku (Fear)**, **Tıksıntı (Disgust)** ve **Küçümseme (Contempt)** sınıflarında başarısız olmuştur. (sınıf dengesizliği ve görsel benzerlik).

Bazı Sınıflar Neden Başarısız Oldu? (Komite Savunması)

- **1. Veri dengesizliği (Class Imbalance)**
- Neutral ve Happiness binlerce örnek
- Contempt ve Disgust çok az örnek
- CNN, istatistiksel olarak **çoğunluk sınıflarına** yönelir.

2 .Görsel benzerlik

- Fear ↔ Surprise
- Disgust ↔ Anger

Tek karelik görüntü bu ayrimı yapmak için yetersizdir.

. ۳ Mikro ifadeler

- Contempt çok küçük kas hareketlerine dayanır
- ΣΛΧΣΛçözünürlükte bu bilgi kaybolur

. ۴ Etiket gürültüsü

FER ۱۰۰ etiketleri %100 temiz değildir.

Sonuçlar

48×48 gri seviye yüz görüntülerinden, sadece ham piksel verisi kullanarak 8 sınıfı duygusal tanıma gerçekleştirildi.

Sonuç olarak, Fear, Disgust ve Contempt sınıflarındaki başarısızlık, veri dengesizliği, düşük çözünürlük, duygular arası görsel benzerlik ve model kapasite sınırlamalarının birleşik etkisinden kaynaklanmaktadır. Bu sınıflar için daha gelişmiş mimariler ve veri stratejileri gereklidir.

Bu sonuçlar, ham piksel tabanlı basit bir CNN ile bile duyguları makul doğrulukta tanıyalabileceğimizi gösteriyor

Örnek Tahminler gradcam uyguladıktan sonra// Grad-CAM Görsel Yorumu

Kırmızı bölgeler → karar için kritik

Mavi bölgeler → önemsiz



Örnek:

- Happiness → ağız , Anger → kaşlar ve gözler , Surprise → göz ve ağız
- Bu, modelin **anlamsız alanlara değil**, yüze odaklandığını kanıtlar.

Gelecek Çalışmalar

- **Akademik Sonuç**

Az temsil edilen sınıflardaki başarısızlık, modelin zayıflığını değil, tek-kare yüz ifadesi tanımanın veri dengesizliği ve görsel belirsizlik altındaki temel sınırlarını göstermektedir.

- Class imbalance için class_weight, focal loss
- Overfitting için EarlyStopping
- Transfer learning (MobileNet, ResNet)



Önerilen İyileştirme Stratejileri:

Bu sınıflardaki başarısızlığı azaltmak için aşağıdaki yöntemler önerilmektedir:

• **Sınıf dengesizliği için:** Class-weight veya Focal Loss kullanımı

• **Overfitting'i azaltmak için:**

EarlyStopping, daha fazla Dropout, daha az epoch.

• **Veri artırma (augmentation) Daha güçlü modeller:**

özellikle Fear ve Disgust sınıfları için hedefli artırma Daha derin CNN veya transfer learning (ResNet, EfficientNet)

• **Yüksek çözünürlüklü ve RGB veri kullanımı**

• **Ek veri setleri:** RAF-DB, AffectNet -----gerçek hayatı daha yakın yüz ifadeleri

Teşekkürler

Dinlediğiniz için teşekkür ederim.