UZAMSAL-ZAMANSAL VİDEO TAHMİNİ

HAZIRLAYAN: SAMET KARTAL
TARİH: 17.12.2025
ÖZET: DERİN ÖĞRENME VE CONVLSTM MİMARİSİ KULLANILARAK, GEÇMİŞ VİDEO KARELERİNDEKİ UZAMSAL VE ZAMANSAL İLİŞKİLERİN ÖĞRENİLMESİ VE GELECEKTEKİ KARELERİN YÜKSEK DOĞRULUKLA TAHMİN EDİLMESİ.

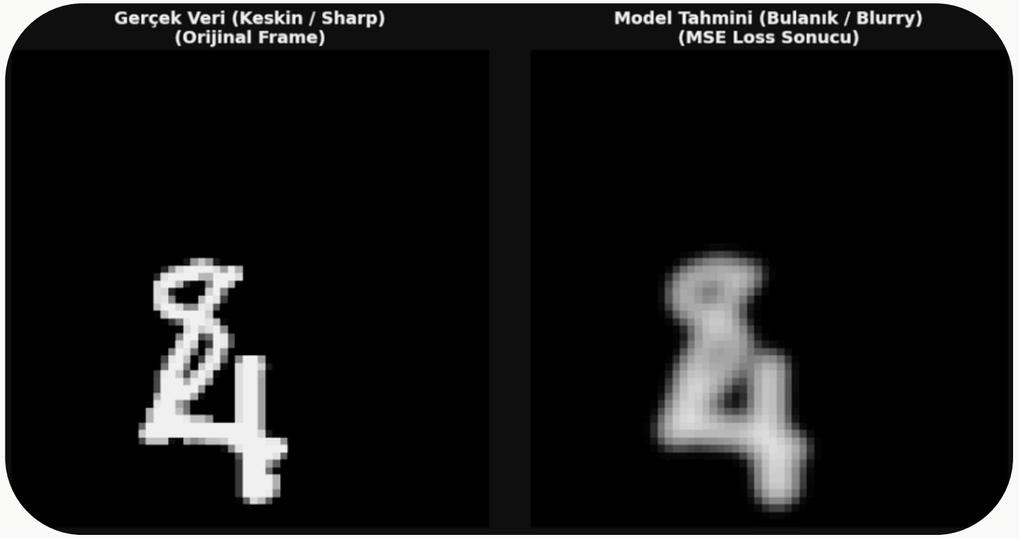
PROJE HEDEFİ VE PROBLEM TANIMI

Video Anahtar Noktalarının Fiziğini Çözmek

Zorluk: Video tahmini, yalnızca piksellerin ara değerlerini doldurma işlemi değildir. Başarılı bir tahmin için modelin hareket, çarpışma ve deformasyon gibi altta yatan fiziksel kuralları öğrenmesi gerekir .

Mevcut Sınırlamalar: Standart modeller genellikle hatayı en aza indirmek için piksellerin ortalamasını alır. Bu yaklaşım ise ne yazık ki bulanık ve net olmayan görüntülerin oluşmasına neden olur .

Hedefimiz: Geliştirdiğimiz hibrit derin öğrenme modeli ile hem görüntünün yapısal bütünlüğünü korumayı hem de zaman içindeki hareketin akıcılığını sağlamayı amaçlıyoruz .



UZAMSAL VE ZAMANSAL KAVRAMLARI

Uzamsal (Spatial) Boyut: Bu kavram, videonun tek bir karesindeki görsel bilgiyi ifade eder. Modelin "Ne görüyorum?" sorusuna yanıt aradığı; şekillerin, dokuların ve nesnelerin algılandığı boyuttur.

Zamansal (Temporal) Boyut: Bu kavram, kareler arasındaki zamanla değişen ilişkiyi tanımlar. Modelin "Nesne nasıl hareket ediyor?" sorusuna odaklandığı; hız, yön ve ivme gibi bilgilerin işlendiği süreçtir.

Sentez: Projemiz bu iki boyutu birleştirerek, nesnelerin hem ne olduğunu hem de nereye gittiğini aynı anda öğrenen bir yapı sunar.

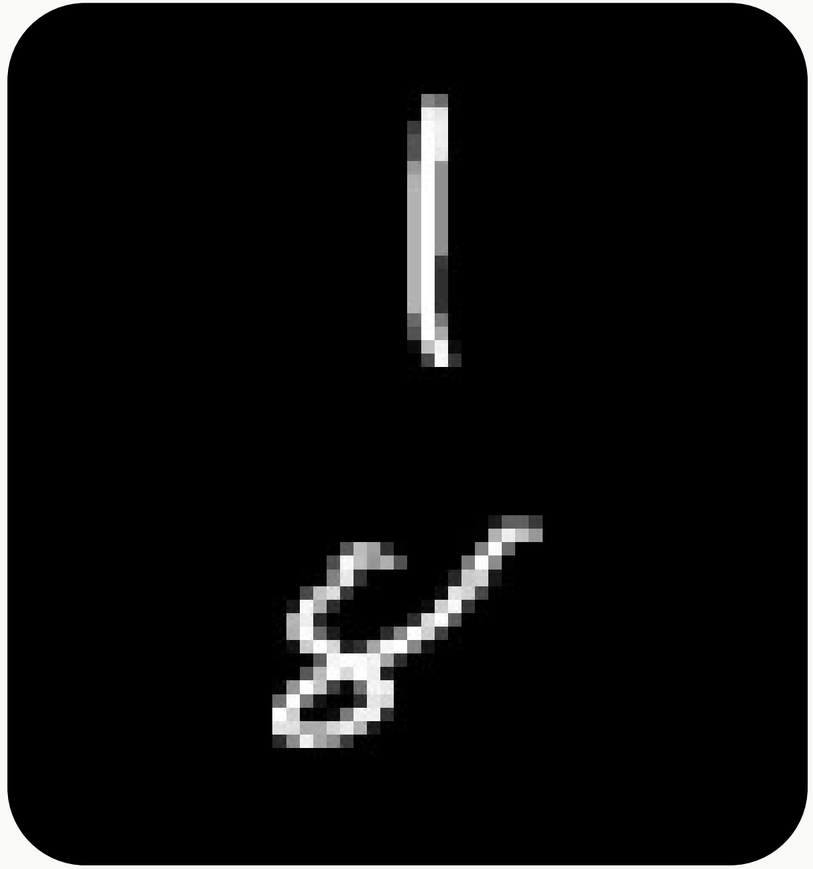
VERİ SETİ

Moving MNIST

Veri Kümesi Yapısı: Toplamda **10.000 adet** video dizimiz var. Her bir video tam **20 kareden** oluşmaktadır.

Öğrenme Mantığı (Next-Frame Prediction): Modelimize "geleceği anlık olarak tahmin etmeyi" öğretiyoruz. 20 karelik bir videoyu şöyle kullanıyoruz:

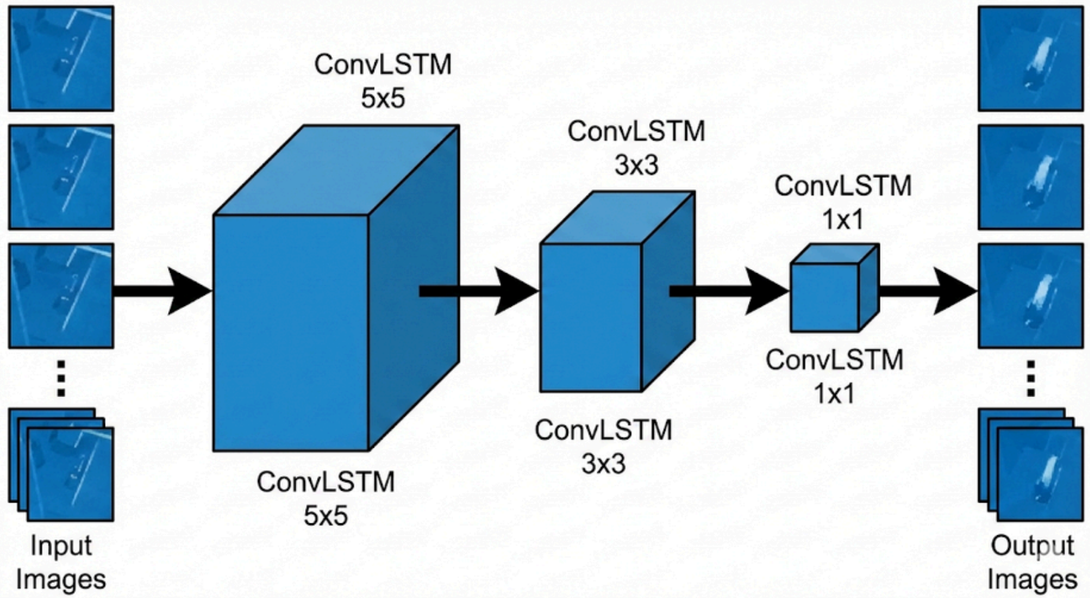
1. **Girdi (Input):** İlk 19 kare (0. indexten 18. indexe kadar).
2. **Hedef (Target):** Bir adım kaydırılmış son 19 kare (1. indexten 19. indexe kadar).



MODEL MİMARİSİ

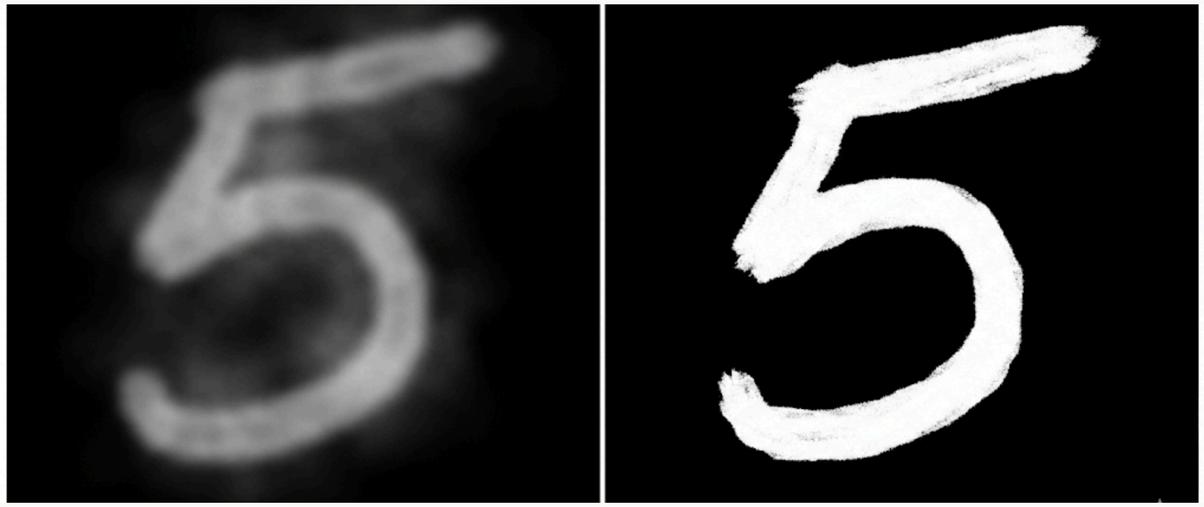
Klasik Encoder-Decoder yapılarının aksine, bu modelde görüntüyü sıkıştırıp küçültmüyoruz. Veriyi her katmanda orijinal çözünürlüğünde (64x64) tutarak, en ufak bir detay kaybını bile engellenir.

- **Hiyerarşik Öğrenme Stratejisi:** Modelimiz 3 aşamalı özel bir filtreleme uygular:
- **5x5 Filtre (Geniş Açı):** Büyük hareketleri ve rakamların genel gidişatını yakalar.
- **3x3 Filtre (Odaklanma)**: Rakamların şeklini ve kenar detaylarını netleştirir.
- **1x1 Filtre (Karar Anı):** Tüm özellikleri birleştirerek son görüntüyü oluşturur.
- **Conv3D Çıktı:** Son aşamada zaman boyutunda (zamansal) bir yumuşatma yaparak videodaki titremeleri (flickering) ortadan kaldırır.



KESKİNLİK VE SEÇİCİ ÖĞRENME

- **LSTM Hafızası (Gating):** Modelimizde harici bir "Attention" katmanı yoktur; bunun yerine LSTM hücrelerinin kendi içindeki "Unutma Kapıları" (Forget Gates) çalışır. Bu kapılar, gereksiz arka plan bilgisini silip sadece hareketli rakamları hafızada tutar.
- **İnsan Gözü Gibi Görme (Perceptual Loss):** Modelimiz sadece pikselleri eşleştirmez; VGG-19 ağı üzerinden görüntünün "içeriğini" ve "anlamını" algılar. Bu sayede insan gözüne hoş gelen doğal görüntüler üretir.
- **Keskin Kenarlar (Gradient Difference Loss):** Siyah zemin üzerindeki beyaz rakamların etrafında oluşan "hale" etkisini ve bulanıklığı cezalandırır. Model, keskin sınırlar çizmeye zorlanır.



KARŞILAŞILAN ZORLUKLAR VE ÇÖZÜMLER

1

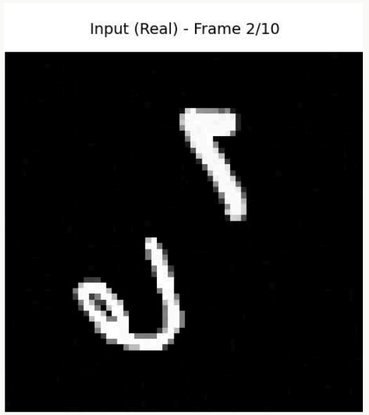
Görüntü Bozulmaları (Artifacts): Rakamlar üzerinde solucan benzeri desenler oluştuğunda, VGG ağındaki algısal kayıp odağımızı dokudan ziyade şekli algılayan katmanlara kaydırarak bu sorunu çözdük .

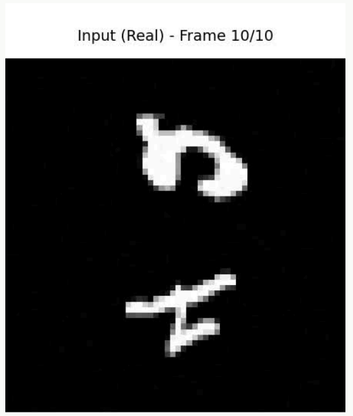
2

Hayalet Görüntü (Ghosting): Hızlı hareketlerde oluşan çift görme sorununu, çıktı aşamasında 3 boyutlu evrişim kullanarak ve zaman ekseninde pürüzsüzlüğü zorlayarak giderdik .

3

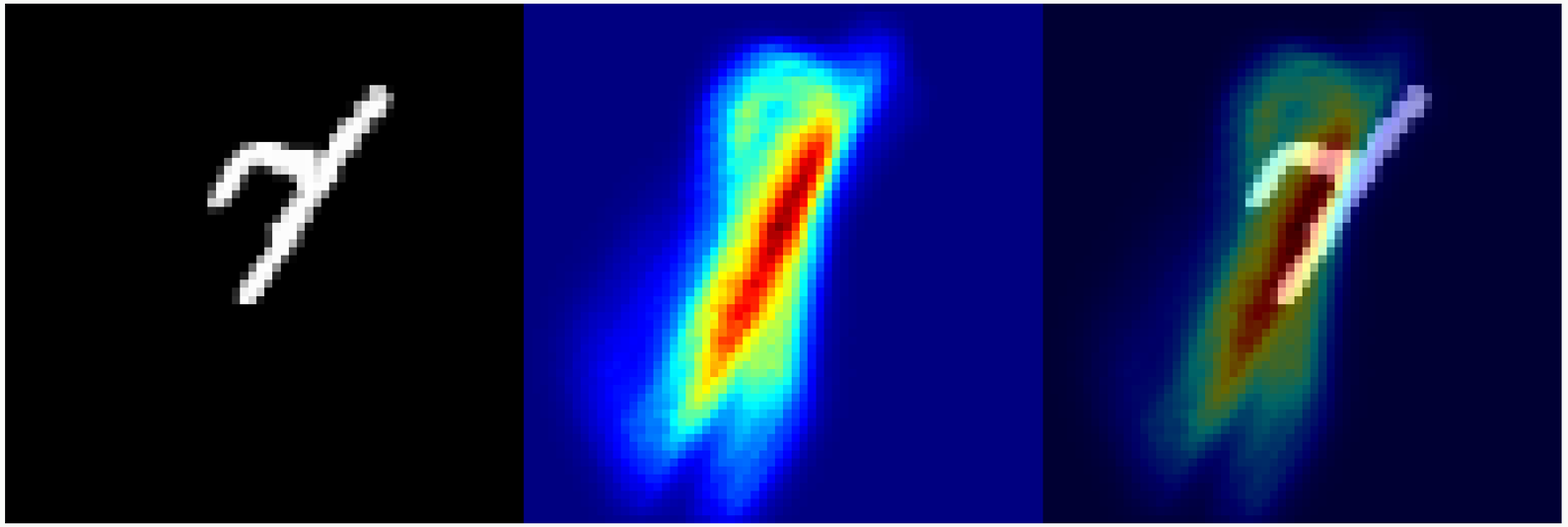
Öğrenmenin Durması (Plateauing): Öğrenme sürecinin tıkandığı noktalarda, hiperparametreleri optimize ederek ve gradyan güncellemelerini sınırlayarak (clipping) eğitimin kararlı bir şekilde devam etmesini sağladık .





BİLEŞENLERİN ETKİSİ

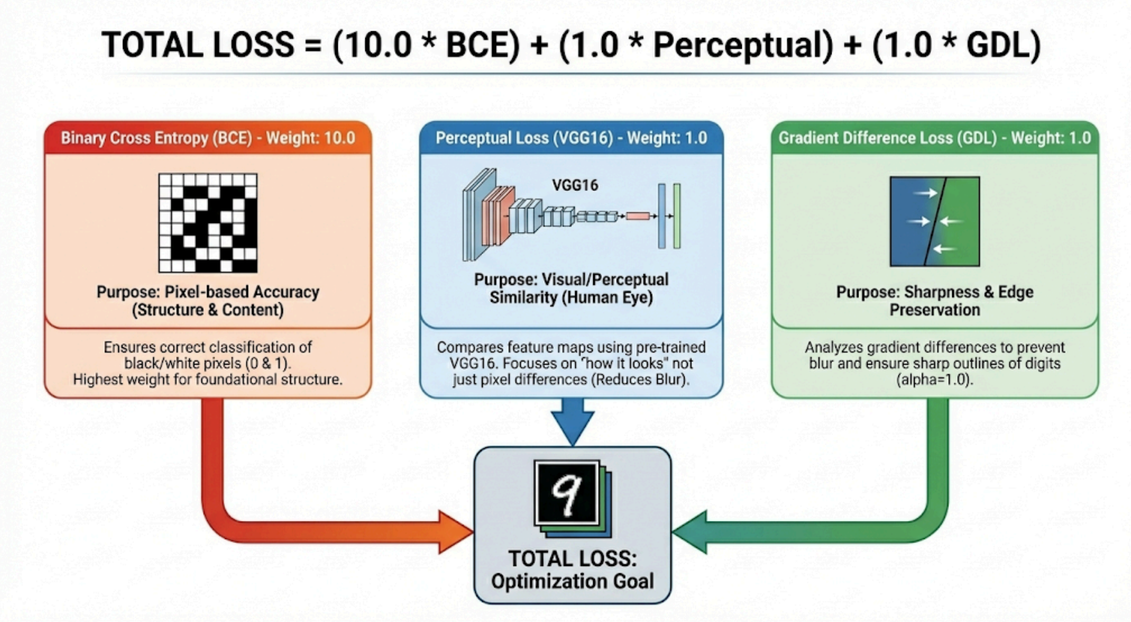
- **Bileşenlerin Uyumu:** Eğer LSTM'in "Unutma Kapıları" düzgün çalışmasaydı, model hızlı hareket eden rakamları takip edemezdi. Benzer şekilde, algısal kayıp fonksiyonu (Perceptual Loss) kullanılmasaydı, elde ettiğimiz görüntü doğru yerde ama hayalet gibi silik olurdu .
- **Aktivasyon Haritası:** Oluşturduğumuz ısı haritaları, modelin dikkatinin tam olarak hareketli rakamlar üzerinde yoğunlaştığını ve siyah arka planı başarıyla görmezden geldiğini kanıtlamaktadır .



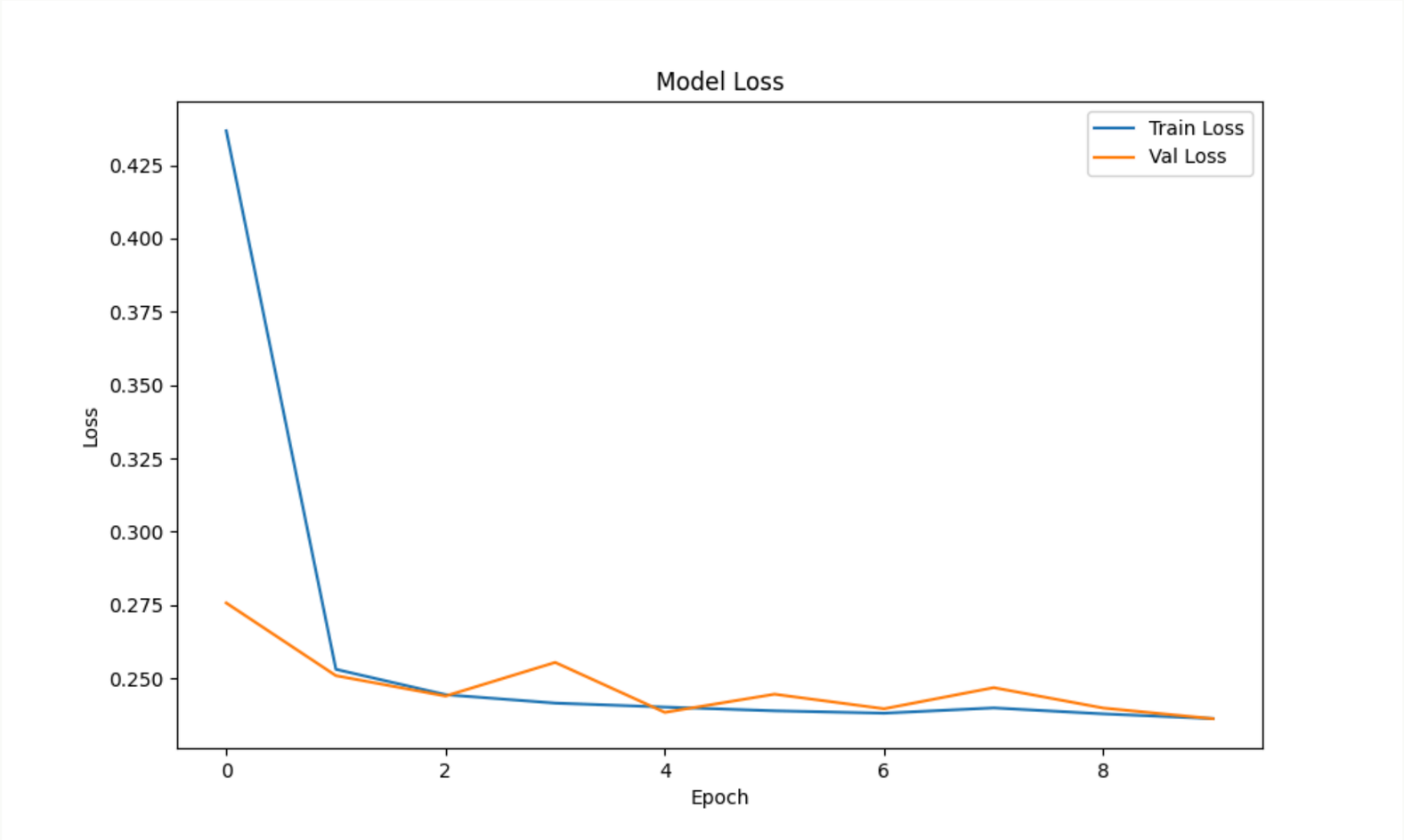
KULLANILAN LOSS YÖNTEMLERİ

Modelin hem piksel doğruluğunu hem de görsel kalitesini maksimize etmek için üç farklı loss fonksiyonunun kombinasyonu kullanılmıştır:

1. **Binary Cross Entropy (BCE)** - Ağırlık: 10.0
   - Amaç: Piksel bazlı doğruluk.
   - Açıklama: Siyah-beyaz piksellerin (0 ve 1) doğru sınıflandırılmasını sağlar. Görüntünün temel yapısının ve içeriğinin doğru tahmin edilmesi için en yüksek ağırlık (10.0) bu fonksiyona verilmiştir.
2. **Perceptual Loss (VGG16)** - Ağırlık: 1.0
   - Amaç: Görsel/Algısal benzerlik.
   - Açıklama: Önceden eğitilmiş VGG16 ağı kullanılarak, gerçek ve tahmin edilen görüntülerin öznitelik haritaları (feature maps) karşılaştırılır. Bu sayede model, sadece piksel farklarına değil, görüntünün "insan gözüne nasıl göründüğüne" odaklanır (Bulanıklığı azaltır).
3. **Gradient Difference Loss (GDL)** - Ağırlık: 1.0
   - Amaç: Keskinlik ve Kenar Koruma.
   - Açıklama: Görüntüdeki gradyan farklarını (kenar geçişlerini) analiz eder. Tahmin edilen karelerdeki bulanıklığı önlemek ve rakamların hatlarının daha keskin olmasını sağlamak için kullanılır (alpha=1.0).



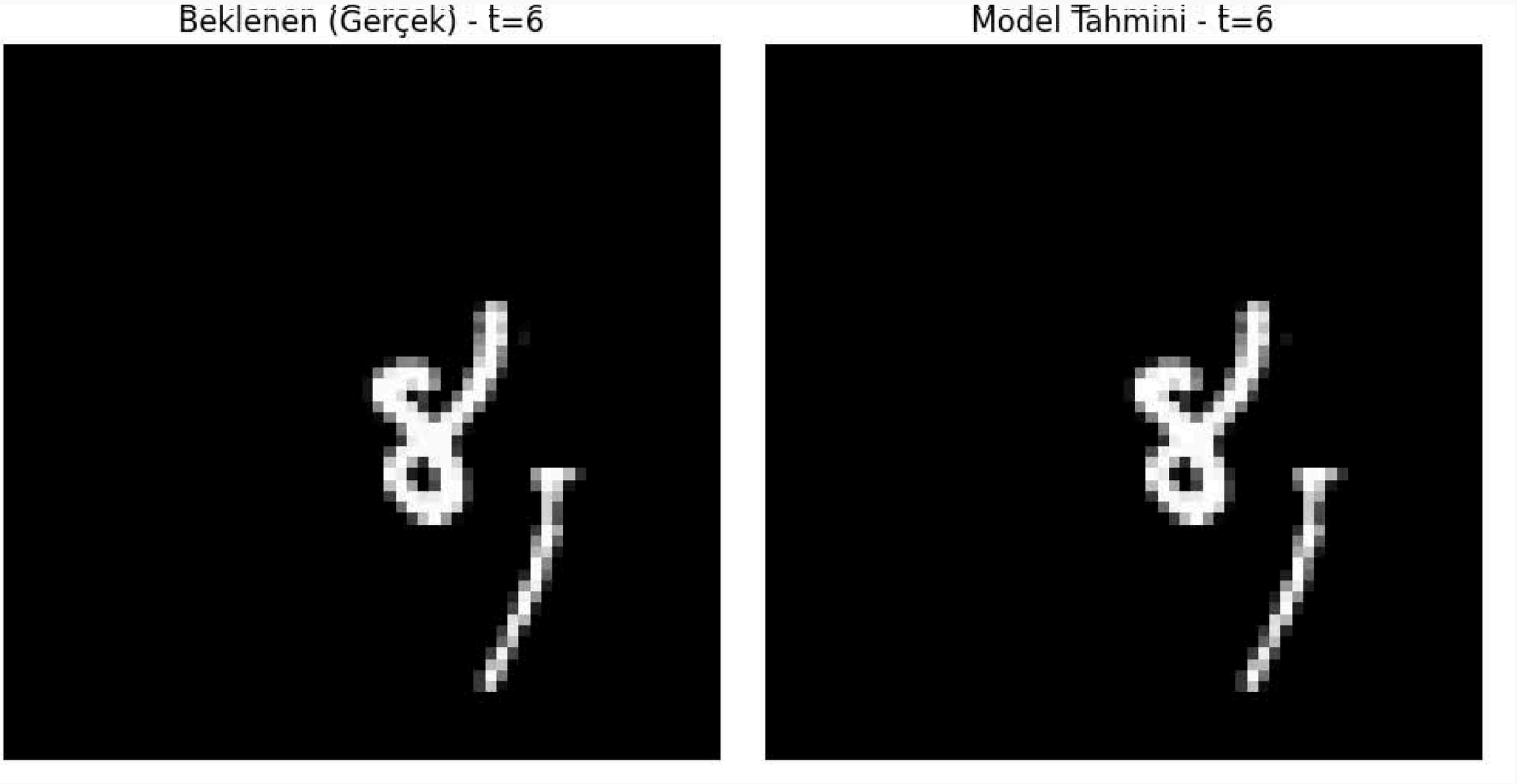
LOSS ANALİZİ



NİHAİ SONUÇLAR

Özet Değerlendirme: Modelimiz, tüm değerlendirme metriklerinde standart referans değerlerin (Benchmark) üzerine çıkarak, hem yapısal hem de piksel bazında yüksek doğruluk sağlamıştır.

PSNR (Peak Signal-to-Noise Ratio)	46.22 dB	40 dB üzeri değerler, insan gözüyle orijinal görüntüden ayırt edilemeyecek kadar yüksek kalite üretildiğini gösterir.
SSIM (Structural Similarity Index)	0.990	Yapısal bütünlük çok yüksek. 1.0'a çok yakın olması, dokuların, kenarların ve şekillerin neredeyse tam olarak korunduğu anlamına gelir.
MSE (Mean Squared Error)	0.08	Çok düşük hata. Piksel başına hata neredeyse yok denecek kadar azdır; modelin tahmini orijinale çok yakın.
MAE (Mean Absolute Error)	0.026	Düşük ortalama sapma. Piksel bazında ortalama fark oldukça küçüktür; bu da global hatanın minimal olduğunu gösterir.


Beklenen (Gerçek) - t=6
Model Tahmini - t=6

TEŞEKKÜRLER