

GAZİ ÜNİVERSİTESİ MÜHENDİSLİK FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ



TÜRKÇE İŞARET DİLİNİN YAZIYA ÇEVİRİLMESİ
FİNAL RAPORU

Aylin AYGÜL 201180060

Aycan KAYNAKCI 191180052

BM466 PERSEPTRON AĞLAR VE UYGULAMALARI (TEK. SEÇ.

Dr. Öğr. Üyesi CEREN GÜZEL TURHAN

2.07.2024

İçindekiler

ŞEKİLLER	iii
1. Giriş.....	1
2. Proje Amacı.....	1
3. Kullanılan Teknolojiler	1
4. Veri Kümesi	2
5. Model Mimarisi.....	4
Eğitim	5
Sonuçlar.....	5
REFERANSLAR	6

ŞEKİLLER

Şekil 1 Videolardan alınan bir görsel	3
---------------------------------------------	---

1. Giriş

Turkish Sign Language Recognazition projesi, görüntülerden işaret dilini tanıyabilen ve bunu metne dönüştürebilen bir sistem geliştirmeyi amaçlamaktadır. Bu sistem, derin öğrenme tekniklerinden yararlanarak bir veri kümesindeki görüntüleri işleyip işaret dilini metne çevirir. Bu projede CNN-LSTM mimarisi kullanılmıştır.

2. Proje Amacı

Projenin ana amacı, işaret dili tanıma sistemi geliştirmektir. Bu sistem, CNN ve LSTM katmanlarını birleştirerek görüntüleri işler ve karşılık gelen işaret dilini metin formatında çıktılar olarak verir.

3. Kullanılan Teknolojiler

Makine Öğrenmesi

CNN-LSTM Mimarisi: Görüntü tanıma ve sıralı veri işleme için Convolutional Neural Network (CNN) ve Long Short-Term Memory (LSTM) katmanları kullanılmıştır.

Veri İşleme ve Görselleştirme

NumPy: Sayısal hesaplamalar ve çok boyutlu veri dizileri için kullanılmıştır.

Matplotlib: Model eğitim aşamasındaki desen ve sonuçları anlamak için veri görselleştirme ve çizim yapma işlemlerinde kullanılmıştır.

Kodlama Ortamı

Python: Geniş kütüphaneleri ve topluluk desteği nedeniyle projenin uygulanması için kullanılan ana programlama dili olmuştur.

Diğer Araçlar

PyTorch: CNN-LSTM modelinin oluşturulması ve eğitilmesi için kullanılan bir derin öğrenme framework'üdür.

Pandas: Büyük veri setlerini manipüle etmek ve analiz etmek için gereken veri yapıları ve işlevleri sağlar.

Model Eğitimi

Google Colab: Derin öğrenme modellerinin verimli bir şekilde eğitilmesi için güçlü GPU'lara ücretsiz erişim sağlayan bir çevrimiçi platformdur.

4. Veri Kümesi

Bu projede kullanılan veri kümesi Ankara Üniversitesi Türk İşaret Dili Veri Kümesi (AUTSL)'dir. Bu veri kümesi, izole edilmiş Türk işaret videolarının büyük ölçekli, çok modlu bir koleksiyonudur.

Veri Kümesinin Temel Özellikleri

İşaret Sayısı: 226 işaret

İşaretçi Sayısı: 43 farklı birey

Toplam Video Örneği: 38.336 video örneği



Şekil 1 Videolardan alınan bir görsel

Veri Önışleme

1. Veri Kumesinin Yükleilmesi: AUTSL veri kümesi, PyTorch'un veri yükleme yardımcı programları kullanılarak yüklenir.
2. Veri Artırma: Dönme, çevirme ve ölçekleme gibi teknikler uygulanarak veri kümesi artırılır ve modelin genelleme yeteneđi geliştirilir.
3. Normalizasyon: Görüntüler, modelin tutarlı giriş sağlaması için normalize edilir.

5. Model Mimarisi

1. CNN-LSTM Mimarisi: Bu mimari, derin öğrenmede görüntü özelliklerini çıkartmak için CNN katmanlarını, zaman serisi veri analizi için LSTM katmanlarını kullanır. Aşağıda modelin yapılandırması verilmiştir:

CNN Encoder: Görüntü verilerini işlemek için evrişimli sinir ağı.

LSTM Decoder: Evrişimli sinir ağı tarafından çıkarılan özellikleri işlemek ve metin çıktısı üretmek için uzun-kısa süreli bellek ağı.

Hyperparametreler:

- Evrişimli Sinir Ağı Katmanları: 6
- LSTM Katmanları: 1
- LSTM Gizli Boyut: 512
- Dropout Oranı: 0.8
- Batch Normalization Kullanımı: Evet
- Yönlendirilebilir (Bidirectional): Evet
- Dikkat Mekanizması Kullanımı: Evet

```
model = CNN_LSTM(
    u_len_label,
    latent_size=512,
    n_cnn_layers=6,
    n_rnn_layers=1,
    n_rnn_hidden_dim=512,
    cnn_bn=True,
    bidirectional=True,
    dropout_rate=0.8,
    attention=True
)
summary(model, input_size=(4, 30, 3, 256, 256))
```

2. Transfer Öğrenme: Önceden eğitilmiş CNN katmanları kullanılabilir ve ardından LSTM katmanlarıyla birlikte ince ayar yapılabilir.

Eğitim

1. Kayıp Fonksiyonu: Modelin performansını ölçmek için çapraz entropi kaybı kullanılır.
2. Optimizatör: Model ağırlıklarını kayıba göre güncellemek için Adam optimizatörü kullanılır.
3. Eğitim Süreci: Model belirli bir epoch sayısı boyunca eğitilir ve her adımda doğrulama yapılır, böylece performans izlenir ve aşırı uyumdan kaçınılır.

Sonuçlar

Bu proje kapsamında, Türk İşaret Dilini (TİD) tanıyan ve metne çeviren bir sistem geliştirilmiştir. CNN-LSTM mimarisi kullanılarak, görüntü işleme ve zaman serisi verileri ile çalışabilen bir model oluşturulmuştur. Eğitim ve doğrulama süreçleri boyunca modelin performansı izlenmiş ve nihai sonuçlar, modelin yüksek doğruluk oranları ile işaret dilini metne dönüştürebildiğini göstermiştir.

Modelin başarısını daha da artırmak amacıyla veri artırma ve transfer öğrenme tekniklerinden yararlanılmıştır. Bununla birlikte, modelin genel performansını değerlendirmek için çeşitli metrikler kullanılmış ve sonuçlar grafiklerle desteklenmiştir.

Gelecekte, modelin genelleme yeteneğini artırmak için daha geniş veri kümeleri üzerinde çalışılması planlanmaktadır. Ayrıca, gerçek zamanlı işaret dili tanıma sistemlerinin geliştirilmesi yönünde adımlar atılacaktır.

REFERANSLAR

1. AUTSL (Ankara University Turkish Sign Language Dataset). (n.d.). Retrieved July 2, 2024, from <https://cvml.ankara.edu.tr/datasets/>
2. Bali, M. (2020). *Gesture Recognition*. GitHub. <https://github.com/mayankbali/Gesture-Recognition>
3. Cooper, H., Holt, B., & Bowden, R. (2011). Sign language recognition. In *Visual Analysis of Humans: Looking at People* (pp. 539-562). London: Springer London.
4. Rastgoo, R., Kiani, K., & Escalera, S. (2021). Sign language recognition: A deep survey. *Expert Systems with Applications*, 164, 113794.
5. Rao, G. A., Syamala, K., Kishore, P. V. V., & Sastry, A. S. C. S. (2018, January). Deep convolutional neural networks for sign language recognition. In *2018 conference on signal processing and communication engineering systems (SPACES)* (pp. 194-197). IEEE.
6. Yang, S., & Zhu, Q. (2017, July). Continuous Chinese sign language recognition with CNN-LSTM. In *Ninth international conference on digital image processing (ICDIP 2017)* (Vol. 10420, pp. 83-89). SPIE.
7. Pigou, L., Dieleman, S., Kindermans, P. J., & Schrauwen, B. (2015). *Sign language recognition using convolutional neural networks*. In European Conference on Computer Vision (pp. 572-578). Springer, Cham.
8. Camgoz, N. C., Hadfield, S., Koller, O., Ney, H., & Bowden, R. (2018). *Neural sign language translation*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7784-7793).
9. Molchanov, P., Yang, X., Gupta, S., Kim, K., Tyree, S., & Kautz, J. (2016). *Online detection and classification of dynamic hand gestures with recurrent 3D convolutional neural networks*. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 4207-4215).
10. Kazakos, E., Nagrani, A., Zisserman, A., & Damen, D. (2019). *EPIC-Fusion: Audio-visual temporal binding for egocentric action recognition*. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 5492-5501).