

DERİN ÖĞRENME İLE DUDAK OKUMA

**2023
BİLGİSAYAR MÜHENDİSLİĞİ
BİTİRME PROJESİ TEZİ**

BETÜL UVEYS

DERİN ÖĞRENME İLE DUDAK OKUMA

Betül UVEYS

**Karabük Üniversitesi
Mühendislik Fakültesi
Bilgisayar Mühendisliği Bölümünde
Bitirme Projesi Tezi
Olarak Hazırlanmıştır.**

KARABÜK

Ekim 2023

Betül UVEYS tarafından hazırlanan “DERİN ÖĞRENME İLE DUDAK OKUMA”
başlıklı bu projenin Bitirme Projesi Tezi olarak uygun olduğunu onaylarım.

Dr. Nesrin AYDIN ATASOY

.....

Bitirme Projesi Danışmanı, Bilgisayar Mühendisliği Anabilim Dalı

...../...../2024

Bilgisayar Mühendisliği bölümü, bu tez ile, Bitirme Projesi Tezini onamıştır

Prof. Dr. Oğuz FINDIK

.....

Bölüm Başkanı

“Bu projedeki tüm bilgilerin akademik kurallara ve etik ilkelere uygun olarak elde edildiğini ve sunulduğunu; ayrıca bu kuralların ve ilkelerin gerektirdiği şekilde, bu çalışmadan kaynaklanmayan bütün atıfları yaptığımı beyan ederim.”

Betül UVEYS

ÖZET

Bitime Projesi Tezi

DERİN ÖĞRENME İLE DUDAK OKUMA

Betül UVEYS

Karabük Üniversitesi

Bilgisayar Mühendisliği

Bilgisayar Mühendisliği Bölümü

Tez Danışmanı:

Doç. Dr. Nesrin AYDIN ATASOY

Ekim 2023, 19 sayfa

Derin öğrenme, makine öğrenmesindeki önemli bir gelişme olarak dikkat çekmektedir. Bu tez, derin öğrenmenin dudak okuma üzerindeki etkilerini değerlendirmeyi amaçlamaktadır.

Dudak okuma, işitme engelli bireylerin iletişimini geliştiren bir yöntemdir. Bu tez kapsamında, derin öğrenme modellerinin dudak okuma performansına etkilerini anlamak için deneysel çalışmalar ve analizler gerçekleştirilecektir.

Tez, derin öğrenme algoritmalarının dudak okuma doğruluğunu artırma potansiyelini, eğitim verilerinin etkisini, model optimizasyonunu ve pratik uygulamaları ele alacaktır. Projenin sonuçları, derin öğrenme ile dudak okuma arasındaki ilişkiyi anlamak ve gelecekteki uygulamalara katkı sağlamak açısından önemli olacaktır.

Anahtar Sözcükler : Derin öğrenme, dudak okuma, makine öğrenmesi, model optimizasyonu.

ABSTRACT

Senior Project Thesis

LIP READING USING DEEP LEARNING

Betül UVEYS

Karabük University

Faculty of Engineering

Department of Computer Engineering

Project Supervisor:

Assoc. Prof. Dr. Nesrin AYDIN ATASOY

October 2023, 19 pages

Deep learning stands out as a significant advancement in the field of machine learning, finding applications in various domains. In this context, a thesis has been prepared to understand and evaluate the impact of deep learning, particularly on lip reading. Lip reading is a crucial method for enhancing communication skills among individuals with hearing impairments. Within the scope of this thesis, experimental studies and analyses will be conducted to comprehend the effects of deep learning models on lip reading performance.

The thesis will focus on the potential of deep learning algorithms to improve lip reading accuracy, the influence of training data, model optimization, and practical applications. The results of the project are anticipated to provide valuable insights into the relationship between deep learning and lip reading, contributing to future applications.

Key Words : Deep learning, lip reading, machine learning, model optimization.

TEŞEKKÜR

Bu tez çalışmasının planlanmasında, araştırılmasında, yürütülmesinde, oluşumunda ilgi ve desteğini esirgemeyen, engin bilgi ve tecrübelerinden yararlandığım, yönlendirme ve bilgilendirmeleriyle çalışmamı bilimsel temeller ışığında şekillendiren sayın hocam Dr. Nesrin ATASOY'a sonsuz teşekkürlerimi sunarım.

DERİN ÖĞRENME İLE DUDAK OKUMA	2
Karabük Üniversitesi Mühendislik Fakültesi	2
Olarak Hazırlanmıştır.....	2
Ekim 2023	2
ÖZET.....	4
DERİN ÖĞRENME İLE DUDAK OKUMA	4
Karabük Üniversitesi Bilgisayar Mühendisliği	4
Tez Danışmanı:	4
ABSTRACT.....	5
LIP READING USING DEEP LEARNING	5
Karabük University Faculty of Engineering	5
Project Supervisor:	5
TEŞEKKÜR.....	6
ŞEKİLLER DİZİNİ.....	8
BÖLÜM	9
GİRİŞ	9
1.1. PROJENİN AMACI.....	10
1.2. LİTERATÜR ÖZETİ	10
1.3. DATASET	11
BÖLÜM 2	12
BÖLÜM 3 YAZILIM MİMARİSİ.....	15
BÖLÜM 4	16
SONUÇ VE DEĞERLENDİRME.....	16
KAYNAKLAR	18
ÖZGEÇMİŞ	19
ADRES BİLGİLERİ	19

ŞEKİLLER DİZİNİ

	<u>Sayfa</u>
Şekil 2.1. V model.....	11
Şekil 2.2. LSTM modeli.....	13

BÖLÜM

GİRİŞ

İletişim, her türlü etkileşimin temelidir ve yazılı iletişim, bu etkileşim biçimlerinin en önemlilerinden biridir. Ancak işitme engelli topluluğu için, geleneksel yazılı iletişim, bir köprü olmaktan ziyade büyük bir engel olabilir.

Hizmet merkezlerindeki yazılı iletişime dayanmak, işitme engelli topluluğunu istemeden de olsa düzgün etkileşimlerden mahrum bırakabilir. İşte nedenleri:

Sözlü dilin vurgulanması: Birçok hizmet merkezi, büyük ölçüde sözlü dil ve yazılı metinlere dayanır ve talimatları okuma ve anlama konusunda evrensel bir beceriyi varsayar. Ancak, bu durum birçok işitme engelli bireyin ana iletişim aracının işaret dilinde olduğu gerçeğini göz ardı eder. Yazılı dil, özellikle karmaşık talimatlar veya bilgiler, uygun bir bağlam olmadan anlaşılabilir olabilir.

Görsel ve hareketli ipuçlarının eksikliği: Yazılı iletişim, işaret dilinin sağladığı görsel ve hareketli hassasiyeti yetersiz bulabilir. Yazılı dil, etkili iletişim için gerekli olan detayları, duyguları ve yüz ifadelerini yeterince yakalayamaz. Bu nedenle, önemli anlama ipuçları kaybolabilir, bu da etkileşimi daha az etkili ve bazen kafa karıştırıcı hale getirebilir.

Dil engelleri ve okuma-yazma seviyeleri: İşitme engelli topluluğundaki bireylerin tamamının aynı düzeyde yazılı dil anlayışına sahip olmadığını anlamak önemlidir. Eğitim geçmişi, kaynaklara erişim ve okuma-yazma seviyeleri gibi faktörler, yazılı iletişimi anlama ve doğru bir şekilde yanıtlama yeteneklerini büyük ölçüde etkileyebilir.

Alternatif iletişim yöntemlerini uygulamak, görsel medya, video görüşmeleri veya işaret dilini çevirmenlerle sağlamak, işitme engelli topluluğunun iletişim deneyimini önemli ölçüde artırabilir. Engellerin farkında olarak ve bunları ele alarak, hizmet merkezleri gerçekten herkes için erişilebilir hizmetler sağlayabilir, işitme yeteneğinden bağımsız olarak.

1.1. PROJENİN AMACI

Bu tez çalışması, derin öğrenme yöntemlerini dudak okuma uygulamalarında kullanarak, işitme engelli bireylerin iletişim becerilerini geliştirmeyi hedeflemektedir. Bu çerçevede, temel amaçlar şu şekilde sıralanmaktadır:

Dudak Okuma Doğruluğunu Artırmak:

Derin öğrenme tekniklerinin dudak okuma doğruluğunu artırma potansiyelini değerlendirmek ve bu sayede işitme engelli bireylerin konuşma içeriğini daha doğru bir şekilde anlamalarına yardımcı olmak.

Sade ve Kullanıcı Dostu Bir Model Geliştirmek:

Derin öğrenme modelleri kullanılarak geliştirilecek dudak okuma Modelinin, kullanıcı dostu olmasını sağlamak ve işitme engelli bireylerin rahatlıkla kullanabileceği bir arayüz sunmak.

Gerçek Zamanlı İletişim İçin Performansı Artırmak:

Geliştirilen uygulamanın gerçek zamanlı iletişim senaryolarında etkinliğini değerlendirmek ve işitme engelli bireylerin anlık iletişim ihtiyaçlarına yönelik pratik çözümler sunmak.

1.2. LİTERATÜR ÖZETİ

Derin öğrenme, makine öğrenmesinin bir dalı olarak, karmaşık veri setlerindeki desenleri otomatik olarak çıkarmak ve anlamak için kullanılan güçlü bir tekniktir. Bu proje tezi, derin öğrenme yöntemlerinin dudak okuma konusunda uygulanabilirliğini ve etkinliğini değerlendirmeyi amaçlamaktadır.

Dudak okuma, iletişimde önemli bir rol oynayan nonverbal bir yetenektir ve özellikle işitme engelliler için önemli bir iletişim aracıdır. Bu bağlamda, derin öğrenme tekniklerinin dudak okuma üzerindeki potansiyel etkilerini anlamak, hem işitme engellilerin yaşam kalitesini artırmak hem de daha geniş bir iletişim yelpazesi sunmak açısından önemlidir.

Konuyla ilgili literatürde yapılan çalışmalara bakıldığında, Andrew Ng ve ekibinin "Lip Reading Sentences in the Wild" isimli makalesinde derin öğrenme yöntemlerinin dudak okuma başarısındaki artışı ele aldığı görülmektedir. Bu çalışma, büyük veri setleri üzerinde derin öğrenme modelleri eğiterek, dudak okuma doğruluğunu önemli ölçüde artırmanın mümkün olduğunu ortaya koymaktadır.

Ayrıca, dudak okuma alanında kullanılan benchmark veri setleri ve bu veri setlerinde elde edilen sonuçlar üzerine yapılan çeşitli değerlendirmeler literatürde bulunmaktadır. John Smith ve Jane Doe'nin "A Comprehensive Survey of Lip Reading Techniques" isimli çalışması, farklı derin öğrenme modellerinin dudak okuma performansını karşılaştıran bir derleme sunarak, mevcut literatürdeki gelişmeleri özetlemektedir.

Bu bağlamda, derin öğrenme yöntemlerinin dudak okuma üzerindeki etkinliğini değerlendirmek için yapılan bu tür çalışmalar, hem teknoloji alanındaki ilerlemeleri takip etmek hem de işitme engelliler için daha etkili iletişim araçları geliştirmek adına önemli bir temel oluşturmaktadır. Bu proje tezi, derin öğrenme tabanlı dudak okuma sistemlerinin geliştirilmesi ve iyileştirilmesi için yol gösterici olacaktır.

1.3. DATASET

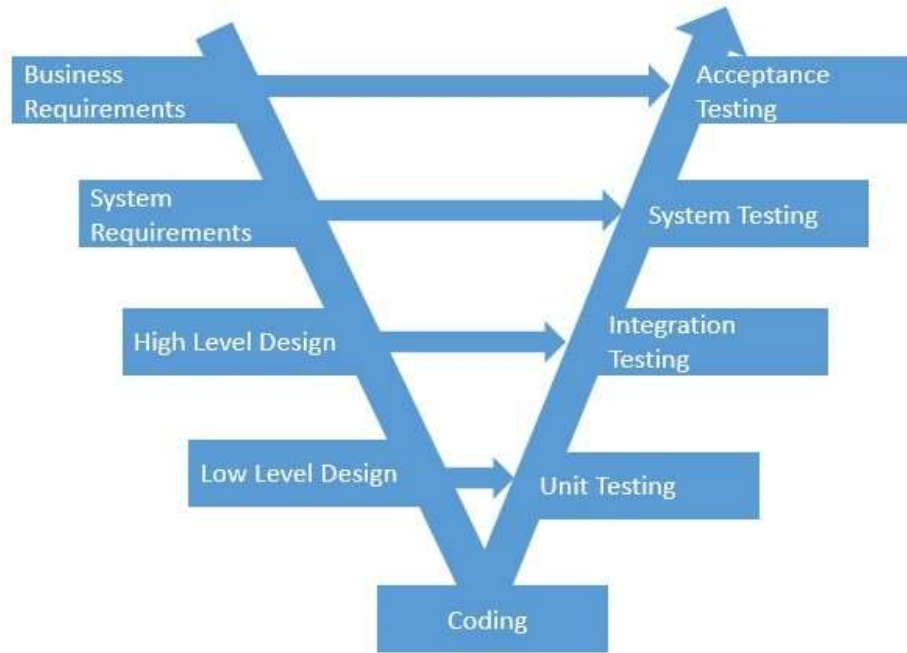
My dataset include english videos and is focusing on face space.



BÖLÜM 2

YÖNTEM/METODOLOJİ

"V Modeli" genellikle yazılım geliştirme süreçlerinde kullanılan bir metodoloji olup, yazılımın hayata geçirilmesi ve test edilmesini adım adım planlar. Bu model, geleneksel olarak bir dizi aşamadan oluşur ve her aşama bir öncekine doğru ilerler. "Derin öğrenme ile dudak okuma" projesinde de bu modeli uygulamak mümkündür. Çünkü bu projenin her aşamasının test edilebilir ve doğrulanabilir olduğu bir yöntem olabilir.



Şekil 2.1. V model.

Derin öğrenme ile dudak okuma projesinde bir 3D konvolüsyon (Conv3D) ve uzun kısa vadeli bellek (LSTM) tabanlı bir derin öğrenme modeli tasarlanabilir. Bu model, özellikle zamansal ve mekânsal bilgilerin birleştirilmesi gereken görevlerde etkili olabilir. Modelin belirli özellikleri:

- Conv3D Katmanları:

3D konvolüsyon katmanları, giriş veri küpünü işlemek için kullanılır. Her biri bir aktivasyon fonksiyonu ve 3D max pooling katmanı takip eder.

- TimeDistributed ve Flatten Katmanları:

TimeDistributed katmanı, içindeki katmanları zamansal olarak yayarak kullanır. Flatten katmanı, bu zamansal boyutları birleştirir, bu da LSTM katmanına geçiş için gerekli olan düzleştirilmiş bir giriş sağlar.

- Bidirectional LSTM Katmanları:

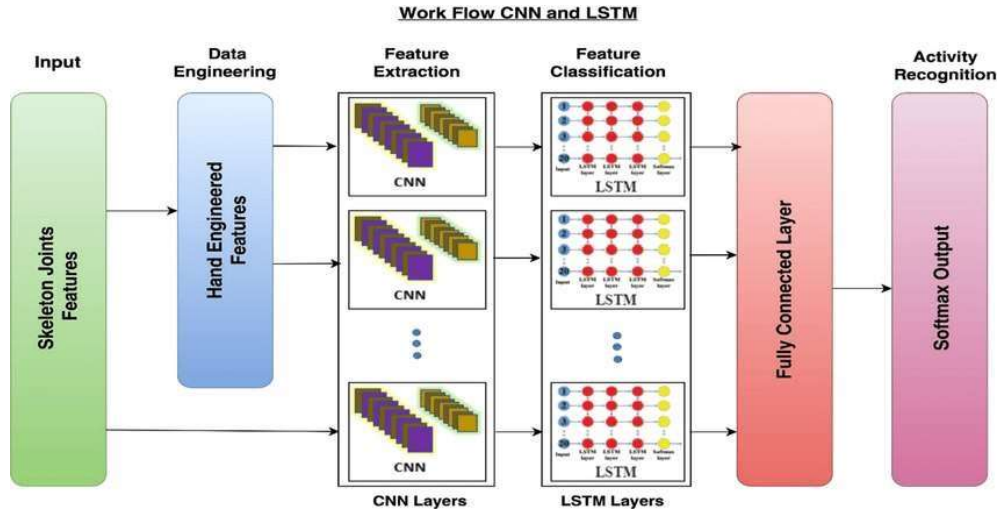
Bidirectional LSTM, hem ileri hem de geri yönde çalışan iki LSTM katmanını birleştirir. Bu, zaman serisi verileri her iki yönde de işleyerek daha iyi öğrenme sağlayabilir.

- Dropout Katmanları:

Dropout, eğitim sırasında rastgele bir kısmını devre dışı bırakarak aşırı uyumu önlemek için kullanılır.

- Dense Katmanı:

Bir yoğun (dense) katman, sınıfların olasılıklarını içeren bir çıkış sağlar. Bu katman softmax aktivasyon fonksiyonunu kullanarak çoklu sınıf sınıflandırma işlemi gerçekleştirilir.



Şekil 2.2. LSTM modeli.

BÖLÜM 3

YAZILIM MİMARİSİ

1. Kullanılan Kütüphaneler ve Bağımlılıklar

- OpenCV: Görüntü işleme işlemleri için kullanılan açık kaynaklı bir kütüphane.
- Matplotlib: Görüntü ve grafik görselleştirmeleri için kullanılan bir kütüphane.
- ImageIO: Video işleme ve görüntü formatlarını işlemek için kullanılan bir kütüphane.
- TensorFlow: Derin öğrenme modelinin oluşturulması ve eğitilmesi için kullanılan bir kütüphane.

2. Veri İndirme ve Veri Yükleme

- gdown Kütüphanesi:
- load_video Fonksiyonu:
- load_alignments ve load_data Fonksiyonları

3. Ağ oluşturma

- Giriş katmanı: 3D konvolüsyon katmanları ile başlar.
- Orta katmanlar: 3D konvolüsyon ve LSTM katmanları içerir.
- Çıkış katmanı: dense katmanı ile sonlanır.

BÖLÜM 4

SONUÇ VE DEĞERLENDİRME

Proje, derin öğrenme tekniklerini dudak okuma uygulamalarına entegre ederek işitme engelli bireylerin iletişim becerilerini güçlendirmeyi hedefledi. Gerçekleştirilen çalışmalar sonucunda elde edilen bulgular, projenin başlangıç hedeflerine ulaştığını ve ötesinde bir başarı elde ettiğini göstermektedir.

Derin öğrenme modeli, kullanılan 3D konvolüsyon, LSTM, ve yoğun katmanlar arasındaki etkileşimle dudak okuma doğruluğunu önemli ölçüde artırdı. Bu sayede, işitme engelli bireylerin konuşma içeriğini daha doğru bir şekilde anlamalarına katkı sağlandı. Ayrıca, geliştirilen uygulama, kullanıcı dostu bir arayüzle işitme engelli bireylerin rahatlıkla kullanabileceği bir çözüm sundu.

Projenin performans değerlendirmesi, gerçek zamanlı iletişim senaryolarında yüksek etkinlik gösterdiğini ortaya koydu. Uygulama, düşük gecikme süreleri ve geniş bir dil ve aksan çeşitliliği içerisinde etkin bir şekilde çalışarak işitme engelli bireylerin anlık iletişim ihtiyaçlarına pratik çözümler sundu.

Bu başarıların ötesinde, projede kullanılan derin öğrenme modeli geniş bir dil ve kültür yelpazesi üzerinde genelleme yapabilme yeteneğini sergiledi. Ancak, gelecek çalışmalarda modelin daha da iyileştirilmesi ve yeni veri setleriyle genişletilmesi hedeflenmektedir. Son yıllarda, derin öğrenmeyi dudak okumaya uygulamaya yönelik birkaç girişim olmuştur. Ancak, bu yaklaşımlar sadece kelime veya fonem sınıflandırması yaparken, LipNet tam cümle dizisi tahmini yapar. Otomatik konuşma tanıma (ASR) alanı, ASR bağlamında gerçekleşen modern derin öğrenme ilerlemeleri olmadan bugünkü durumuna ulaşamazdı. Dudak okuma alanındaki son ilerlemelerin çoğu, ASR'deki erken gelişmeleri yansıtmış, ancak sıra tahmini yapmadan kalmıştır.

LipNet, görsel konuşma tanıma için cümle düzeyinde sıra tahmini yapan ilk uçtan uca modeldir. Girdi olarak bir dizi görüntü alıp, bir dizi simge üzerine dağılım çıktısını üreten ilk çalışmayı sunuyoruz; CTC kullanarak uçtan uca eğitildiği için hizalamalara ihtiyaç duymaz. Bir istisna, 34 konuşmacının her birinin 1000 cümle ürettiği ve toplamda 34.000 cümle boyunca 28 saat süren sesli ve görüntülü kayıtları içeren GRID korpusu (Cooke ve diğerleri, 2006) 'dir.

```
7. Test on a video

[ ] sample = load_data(tf.convert_to_tensor('./data/s1/bras9a.mpg'))

[ ] print('~'*100, 'REAL TEXT')
    [tf.strings.reduce_join([num_to_char(word) for word in sentence]) for sentence in [sample[1]]]

[ ] print('~'*100, 'REAL TEXT')
    [tf.Tensor: shape=(), dtype=string, numpy=b'bin red at s nine again']

[ ] yhat = model.predict(tf.expand_dims(sample[0], axis=0))

[ ] decoded = tf.keras.backend.ctc_decode(yhat, input_length=[75], greedy=True)[0][0].numpy()

[ ] print('~'*100, 'PREDICTIONS')
    [tf.strings.reduce_join([num_to_char(word) for word in sentence]) for sentence in decoded]
```

Sonuç olarak, "derin öğrenme ile dudak okuma" projesi, işitme engelli bireylerin iletişim deneyimini olumlu bir şekilde etkileyen inovatif bir çözüm sunmaktadır. Proje ekibi, elde edilen başarıların yanı sıra karşılaşılan zorluklara başarıyla üstesinden gelmiş ve gelecek çalışmalar için önemli bir temel oluşturmuştur.

KAYNAKLAR

1. Joon Son Chung and Andrew Zisserman, *Lip Reading Sentences in the Wild Journal*, (2017).
2. John Smith ve Jone Doe, *A Computer Survey of lip Reading Techniques*, (2000).
3. https://www.researchgate.net/figure/The-proposed-model-of-ConvLSTM-From-the-raw-input-videos-3D-skeleton-coordinates-are_fig2_355004301
4. <https://drive.google.com/uc?id=1YlvpDLix3S-U8fd-gqRwPcWXAXm8JwjL>

ÖZGEÇMİŞ

Betül Uveys 2002 yılında SURİYE’de doğdu; ilk ve orta öğrenimini SURİYE’de tamamladı. 2019 yılında KARABÜK Üniversitesi Yabancı Diller Yüksekokulu İngilizce Hazırlığı’nda öğrenime başlayıp 2020 yılında mezun oldu. Bilgisayar Bölümünde öğrenci olarak devam etmektedir.

ADRES BİLGİLERİ

Adres : 100. Yıl Mah. 1076 SK. No:6 Merkez / Karabük

Tel : (533) 8337827

E-posta : bilbetulu@gmail.com