KOÇ UNIVERSITY

Duygulanımsal Konuşma ve İşmar Modelleri için Derin Öğrenme

ENGİN ERZİN

Bilgisayar ve Elektrik/Elektronik Mühendisliği Bölümleri, Koç Üniversitesi

Konuşma Planı

- Alana Genel Giriş
  - Davranışsal/Sosyal Sinyal İşleme
- Konuşma ve İşmar Modelleri Çalışmalarımız
- Derin Modellerde Zorluklar
- Aktarmalı Öğrenme
  - Marjinalleştirilmiş Gürültü Giderici Yığın Otokodlayıcı (mGGYO)
  - mGGYO Tabanlı Çalışmalardan Örnekler ve Sonuçlar
- Vargılar

Davranışsal/Sosyal Sinyal İşleme

- Disiplinler arası
  - Mühendislik - Psikoloji – Tıp
- İnsan davranışlarının hesaplamalı analizi
  - Hızlı, tutarlı ve otomatik analiz
- Geniş kapsamlı çalışmalar
  - Farklı kipler
  - Veri madenciliği
  - Zengin uygulamalar

Davranışsal/Sosyal Sinyal İşleme

- İnsan davranışları
  - Zihin-beyin-vücut etkileşimi
  - Yansımaları iletişim, sosyal etkileşim ve kişilik özelliklerinde
  - İçeriğindeki öğeler:
    - Duygu, duygudurum, empati, dikkat, ilgi
  - Çok-kipli sinyal işleme
    - Konuşma – Akustik/Sözel
    - Video, Hareket
    - Bio sensörler

Problemimiz: Konuşma ve İşmar Modeli

- İşmar
  - **Sözlük**: El, göz, kaş ya da başla yapılan ve bir şey anlatmaya çalışan işaret
  - **Bu sunumda**: İçgüdüsel veya planlı el, kol, kafa, yüz hareketleri
- Konuşma ve işmar
  - Zamanda senkron ve birliktedirler
  - Birlikte üretilirler
  - Planlama ve şekillenme duygudurum etkisindedir
- Konuşma ve işmar arasındaki istatistiksel ilintiyi nasıl modelleriz?

Neden Önemli?

- Etkileşim izleme ve analizi
  - İnsan-makine etkileşimi için duygudurum/davranış/dikkat/ilgi izlemek
  - Psikolojik hastalıklar, otizm, ADD tanısı, takibi, tedavisi
- Doğal etkileşimler yaratmak için
  - Konuşma ile sürülen animasyonlar/canlandırmalar
  - İntonasyon/vurgu ve işmar sentezi
  - İnsan-makine etkileşimi, film/animasyon/oyun uygulamaları

Nasıl bir Çözüm?

- Çok-kipli konuşma ve işmar analizi
  - İlinti modelleri
    - Güçlü veya zayıf
    - Yapısal olan veya olmayan
  - Kipler için birimler
    - Eğitmenli veya eğitmensiz kümeleme
  - Zamanda ilinti yapıları
    - Saklı Markov modelleri (HMM)
    - Özyinelemeli ağlar (RNN)
    - Uzun kısa soluklu bellek ağları (LSTM)

Çok-Kipli Konuşma-İşmar Çalışmalarımız

Çok-Kipli Konuşma-İşmar Çalışmalarımız

Konuşma ile sürülen kafa hareketleri (2007)

Konuşmadan duygu sınıflandırma (2011)

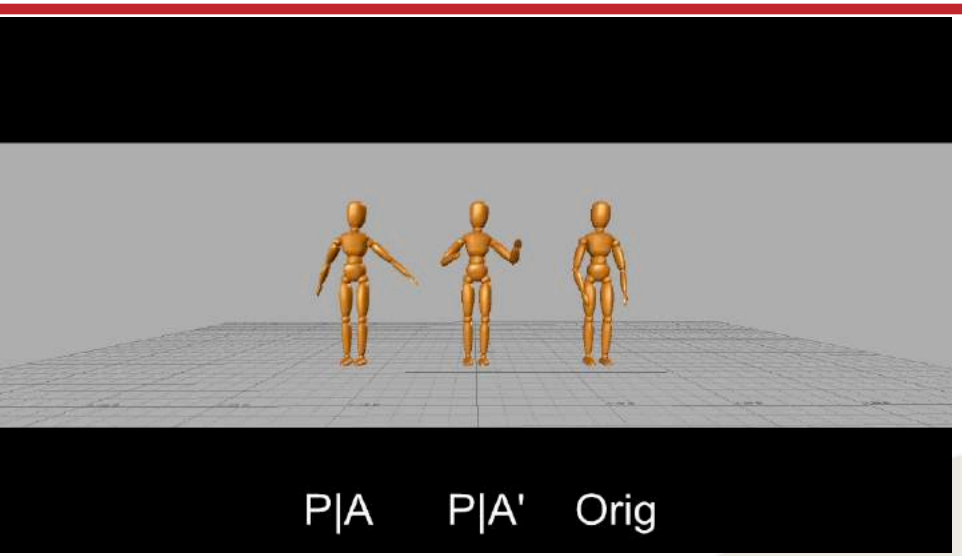
Konuşma ile sürülen üst gövde sentezi (2012)

Konuşmadan duygu takibi (2015)

Duygulanımsal konuşma ve işmar modelleri (2016-18)

- Konuşma ve duygudurum ile sürülen işmar sentezi
- İşitsel-görsel gülme kestirimi
- Konuşma ve duygudurum ile sürülen yüz sentezi

Çok-Kipli Konuşma-İşmar Çalışmalarımız



Konuşma ile sürülen kafa hareketleri (2007)

Konuşmadan duygu sınıflandırma (2011)

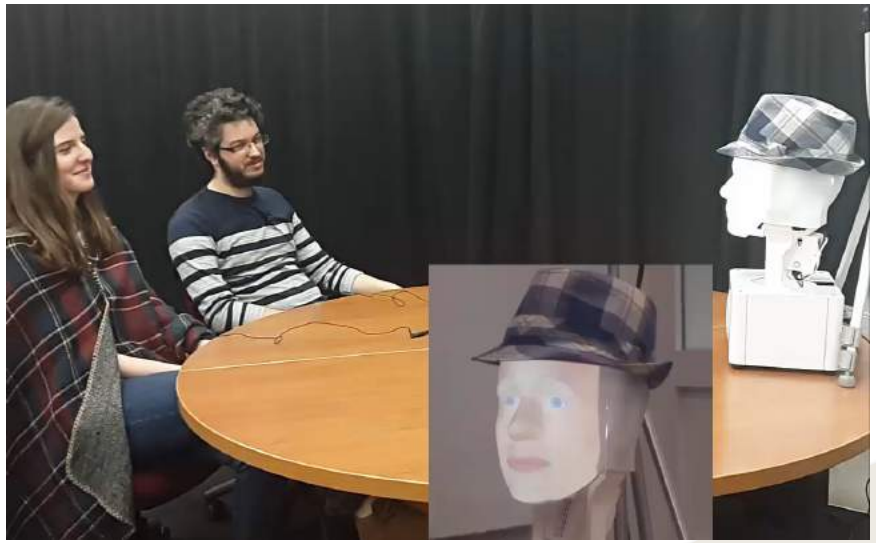
Konuşma ile sürülen üst gövde sentezi (2012)

Konuşmadan duygu takibi (2015)

Duygulanımsal konuşma ve işmar modelleri (2016-18)

- **Konuşma ve duygudurum ile sürülen işmar sentezi**
- İşitsel-görsel gülme kestirimi
- Konuşma ve duygudurum ile sürülen yüz sentezi

Çok-Kipli Konuşma-İşmar Çalışmalarımız



Konuşma ile sürülen kafa hareketleri (2007)

Konuşmadan duygu sınıflandırma (2011)

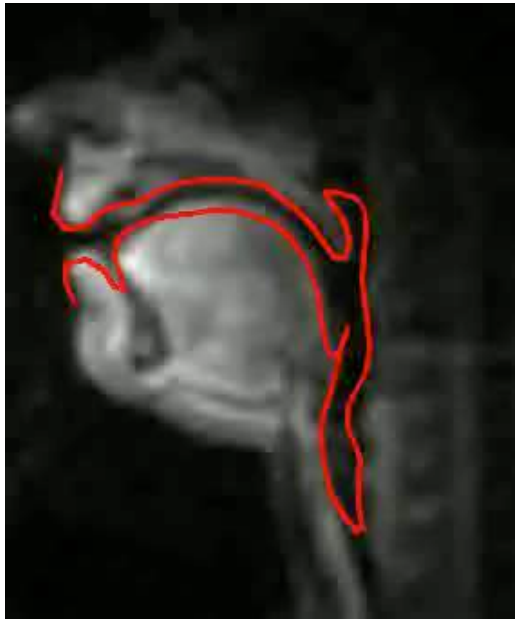
Konuşma ile sürülen üst gövde sentezi (2012)

Konuşmadan duygu takibi (2015)

Duygulanımsal konuşma ve işmar modelleri (2016-18)

- Konuşma ve duygudurum ile sürülen işmar sentezi
- **İşitsel-görsel gülme kestirimi**
- Konuşma ve duygudurum ile sürülen yüz sentezi

Çok-Kipli Konuşma-İşmar Çalışmalarımız



Konuşma ile sürülen kafa hareketleri (2007)

Konuşmadan duygu sınıflandırma (2011)

Konuşma ile sürülen üst gövde sentezi (2012)

Konuşmadan duygu takibi (2015)

Duygulanımsal konuşma ve işmar modelleri (2016-18)

- Konuşma ve duygudurum ile sürülen işmar sentezi
- İşitsel-görsel gülme kestirimi
- Konuşma ve duygudurum ile sürülen yüz sentezi
- **Ses yolu takibi**

Ne Tür Veriler Üzerinde Çalışıyoruz

- İnsan-insan veya insan-robot etkileşim verileri
  - Çok-kipli
    - Konuşma, video, hareket
  - Etiketli
    - Duygudurum sınıflandırılması yapılmış: Mutlu, üzgün, kızgın, vb.
    - Duygudurum etiketlemesi yapılmış: Aktivasyon, Hoşluk, Baskınlık uzaylarında
  - Kaynak zenginliği?
    - Duygulanımsal çok-kipli ve etiketli veri kümeleri limitli!

Derin Modellerde Zorluklar

- Limitli eğitim ve sınama verisi
  - Veri kümeleri arası farklar, değişimler, …
- Olası çözümler
  - Kısıtlı bir alanda kalmak
  - Geniş veri kümeleri oluşturmak
    - Etiketleme zor ve zahmetli!
  - Aktarmalı öğrenme ile daha gürbüz modeller üretmek

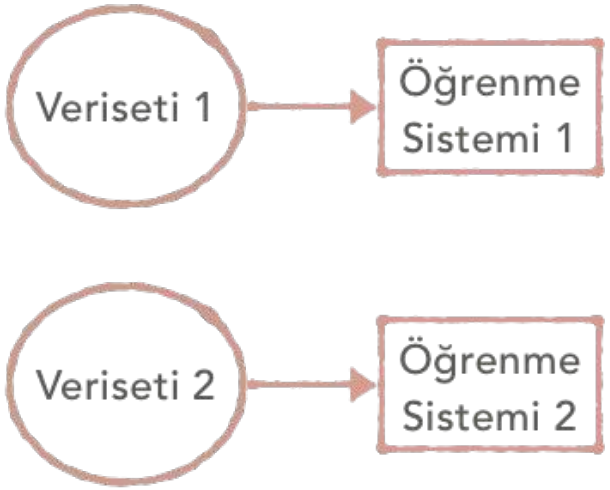
Aktarmalı Öğrenme

- Önceki görevlerde öğrenilen bilgileri yeni görevlere uygulama
  - İnsan zihninin öğrenmesine benzer bir yaklaşım
  - Önceden öğrenilenlerin yeni durumlara aktarımı
  - Yeni görevler önceki deneyimlerle ne kadar ilintili ise, aktarım o kadar kolay olur
    - Gitar çalmak → Bağlama çalmak
    - Araba sürmek → Motosiklet sürmek
    - Matematik → Makine öğrenme

Aktarmalı Öğrenme

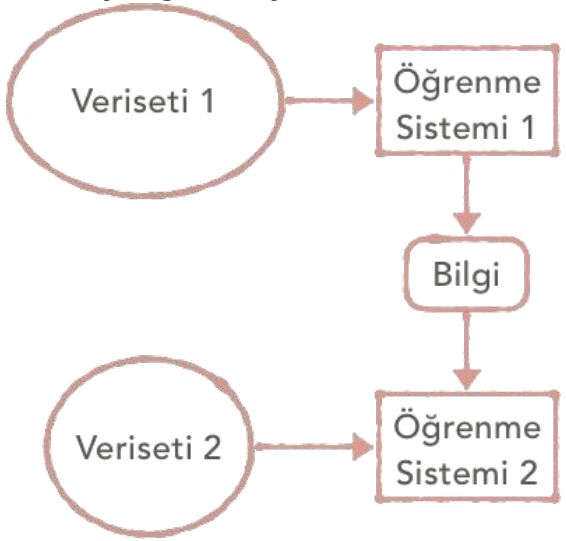
Geleneksel Makine Öğrenme

- Ayrık görevler vardır ve tek görevli öğrenme gerçekleşir
- Bilgi korunmaz veya biriktirilmez
- Diğer görevlerde öğrenilen bilgiler göz önünde bulundurulmaz

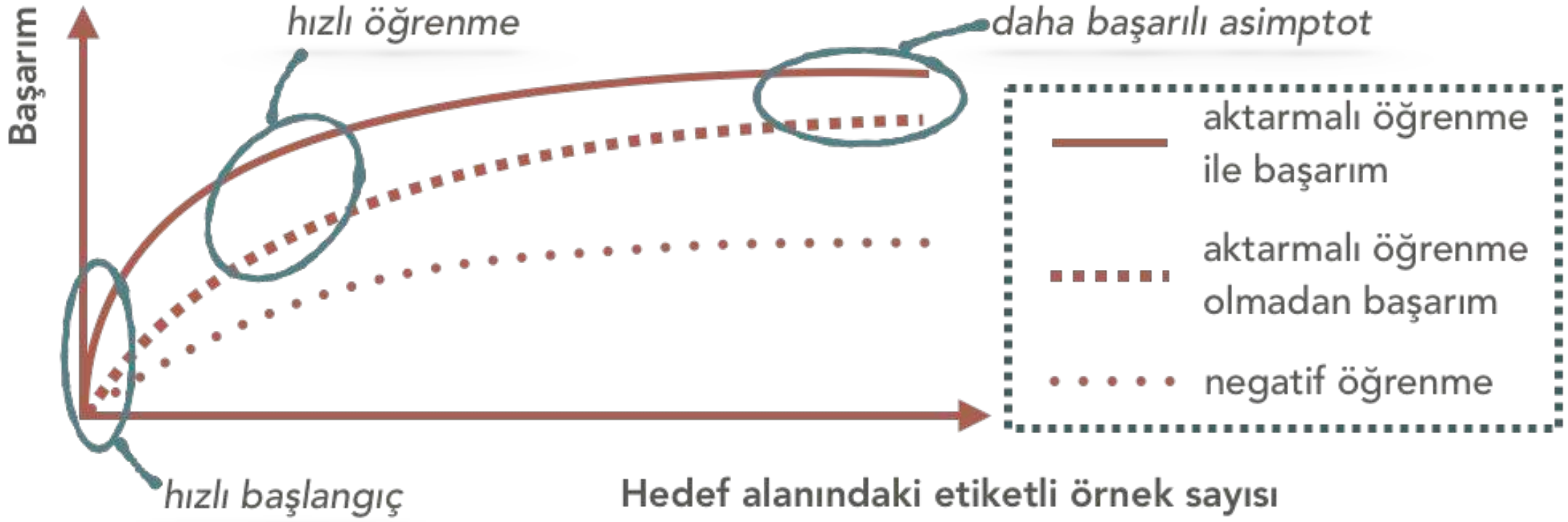


Aktarmalı Öğrenme

- Yeni bir görevin öğrenilmesi daha önceden öğrenilen görevlere dayanır
- Öğrenme süreci daha hızlı, daha doğru olabilir ve/veya daha az eğitim verisine ihtiyaç duyulabilir



Avantajlar ve Limitler



- Öğrenme yapılmadan önceki ilk başarım, daha yüksek olabilir
- Aktarılan bilgi, hedef görevi öğrenmek için gerekli olan süreyi azaltabilir
- Aktarılan bilgiyi kullanarak hedef görevde elde edilebilen başarım artabilir

Aktarmalı Öğrenme Tanım (1)

- Bir alanı (domain) 4 bileşen ile tanımlayabiliriz:

$$D = \{\mathrm{X}, Y, P(X), P(Y|X)\}$$

  - Öznitelik uzayı: X, $X = \{x_1, x_2, \dots, x_n\}$
  - Marjinal dağılım: $P(X)$
  - Etiket uzayı: Y, $Y = \{y_1, y_2, \dots, y_n\}$
  - Koşutlu dağılım / Kestirim fonksiyonu: $P(Y|X)$, $h\colon \mathrm{X} \to \mathrm{Y}$

Aktarmalı Öğrenme Tanım (2)

- Kaynak ve hedef alanları ile bunlara bağlı iki görevi düşünelim,
  - Kaynak alanı:

$$D_S = \{X_S, Y_S, P(X_S), P(Y_S|X_S)\},$$

$$X_S = \{x_{S_1}, \dots, x_{S_n}\}, Y_S = \{y_{S_1}, \dots, y_{S_n}\}$$

  - Hedef alanı:

$$D_T = \{X_T, Y_T, P(X_T), P(Y_T|X_T)\},$$

$$X_T = \{x_{T_1}, \dots, x_{T_m}\}, Y_T = \{y_{T_1}, \dots, y_{T_m}\}$$

Aktarmalı Öğrenme Tanım (3)

- Kaynak/Hedef alanları ve öğrenme görevleri göz önüne alındığında, farklı alanlar veya farklı görevler olabilir:
  - Alan farklılığı

$$X_S \neq X_T \qquad P(X_S) \neq P(X_T)$$

  - Görev farklılığı

$$Y_S \neq Y_T \qquad P(Y_S|X_S) \neq P(Y_T|X_T)$$

- **Aktarmalı öğrenme**: Kaynak bilgisini kullanarak hedef tahminini geliştirmeyi amaçlar

Aktarmalı Öğrenme – Örnek Kanser Verisi

$P(X_S)$

Yaş Sigara

$$X_S = \{x_1^S, x_2^S\}$$

$$X_S \neq X_T$$

$$P(X_S) \neq P(X_T)$$

$P(X_T)$

Yaş Sigara Boy

$$X_T = \{x_1^T, x_2^T, x_3^T\}$$

Aktarmalı Öğrenme – Örnek Kanser Verisi

$P(X_S)$

Yaş Sigara

$$X_S = \{x_1^S, x_2^S\}$$

Kaynak Görevi: Kanser veya değil şeklinde sınıflandırma

$$Y_S \neq Y_T$$

$P(X_T)$

Yaş Sigara Boy

$$X_T = \{x_1^T, x_2^T, x_3^T\}$$

Hedef Görevi: Kanser derecesini 1-2-3 olarak sınıflandırma

Aktarmalı Öğrenme Taksonomi

- Homojen $X_S = X_T$ ve Heterojen $X_S \neq X_T$
- Bunlara ek 4 temel kategori:

1. **Örnek tabanlı aktarım**: Kaynak alanındaki verilerin bir kısmı yeniden ağırlıklandırılarak (reweighting) hedef alanında kullanılır
2. **Öznitelik tabanlı aktarım**: Hedef alanda en uygun öznitelik gösterimini oluşturmak için iki alan arasındaki dağılım farklarını azaltacak bir gizli öznitelik uzayı (latent space) ortaya çıkartılır
3. **Parametrik aktarım**: Kaynak ve hedef görevlerinin bazı parametreleri ya da bunlara ait öncül dağılımlar paylaşılır
4. **İlişki tabanlı aktarım**: Genelde kullanılan uygulamaya özgü kaynak ve hedef verilerindeki benzer ilişkiler kullanılır

Öznitelik Tabanlı Aktarmalı Öğrenme

- Elimizdekiler
  - Etiketli kaynak: $D_S = \{X_S, Y_S\},\ X_S = \{x_{S_1}, \dots, x_{S_{n1}}\}$
  - Etiketsiz hedef: $D_T = \{X_T\},\ X_T = \{x_{T_1}, \dots, x_{T_{n2}}\}$
  - Etiketsiz paralel veri: $D_C = \{X_S^c, X_T^c\} = \{(x_{S_1}^c, x_{T_1}^c), \dots, (x_{S_{nc}}^c, x_{T_{nc}}^c)\}$
- Problem
  - $W_S X_S$ ve $\mathrm{G} \cdot W_T X_T$ gizli ortak değişken uzayı birbirine yaklaştırmak için
  - İzdüşüm matrislerini öğrenmek: W_S ve W_T (homojen öznitelik öğrenme)
  - Dönüşüm matrisini öğrenmek: G (heterojen öznitelik öğrenme)

Homojen Öznitelik Öğrenme

- Marjinalleştirilmiş Gürültü Giderici Yığın Otokodlayıcı (mGGYO)

  (Marginalized Stacked Denoised Autoencoder (mSDA))

  - Bir kodlayıcı ve bir kod çözücüden oluşur
  - Yeniden yapılandırma hatasını en aza indirmek için eğitilir
  - **Yığın mimari**: İlk kodlanmış çıktıyı eğitim verisi olarak gören ikinci bir otokodlayıcı eğitilebilir
  - **Gürültü giderici yığın otokodlayıcı**: Girdi vektörünün stokastik olarak bozulduğu ve modelin onu etkisiz hale getirecek şekilde eğitildiği yapıdır

Homojen Öznitelik Öğrenme – mGGYO (1)

- İzdüşüm matrisi W_S kaynak alanındaki veri için karesel hata maliyetini en aza indirgeyerek öğrenilir

$$\sum_{i=1}^{m} \| X_S - W_S X_S^{(i)} \|^2$$

burada $X_S^{(i)}$ gürültülü kaynak verisini temsil eder.

- Çözüm en küçük karesel hatalar yöntemi ile bulunabilir

$$W_S = PQ^{-1}, \qquad Q = \tilde{X}_S \tilde{X}_S^T \qquad P = \bar{X}_S \tilde{X}_S^T$$

burada $\bar{X}_S = [X_S \cdots X_S]$, $\tilde{X}_S = [X_S^{(1)} \cdots X_S^{(m)}]$

Homojen Öznitelik Öğrenme – mGGYO (2)

- İzdüşüm doğrusal olmayan hiperbolik tanjant ile sıkıştırılır

$$H_S = \tanh(W_S X_S)$$

- Yığın mimarisi ile katmanlar çoğaltılabilir

$$\{W_{S,k}, H_{S,k}\} \qquad k = 1, \dots, K$$

- Öznitelikler üzerinde yanlılık olmadığı durumlarda W_S birim matris olur!

Heterojen Öznitelik Öğrenme – mGGYO

- Öznitelikler arası dönüşüm matrisi G her katmanda heterojen öznitelikler $H_{S,k}^{c}$ ve $H_{T,k}^{c}$ arası

$$\underset{G_k}{argmin}\{\| H_{S,k}^{c} - G_k H_{T,k}^{c} \|^2 + \lambda \| G_k \|^2\}$$

hedef fonksiyonunu en küçükleyerek bulunur.

- Buradan elde edilen çözüm şöyle olur:

$$G_k = (H_{S,k}^{c} {H_{T,k}^{c}}^T)(H_{T,k}^{c} {H_{T,k}^{c}}^T + \lambda I)^{-1}$$

Hibrit Heterojen Aktarmalı Öğrenme

- İlklendir:

$$H_{S,1} = [X_S \; X_S^c], \qquad H_{T,1} = [X_T \; X_T^c]$$

$$\underset{G_1}{argmin}\{\| H_{S,1}^c - G_1 H_{T,1}^c \|^2 + \lambda \| G_1 \|^2\}$$

- Tekrarla $\mathrm{k} = 2, \ldots, K$

$$\{W_{S,k}, H_{S,k}\} = mGGYO(H_{S,k-1})$$

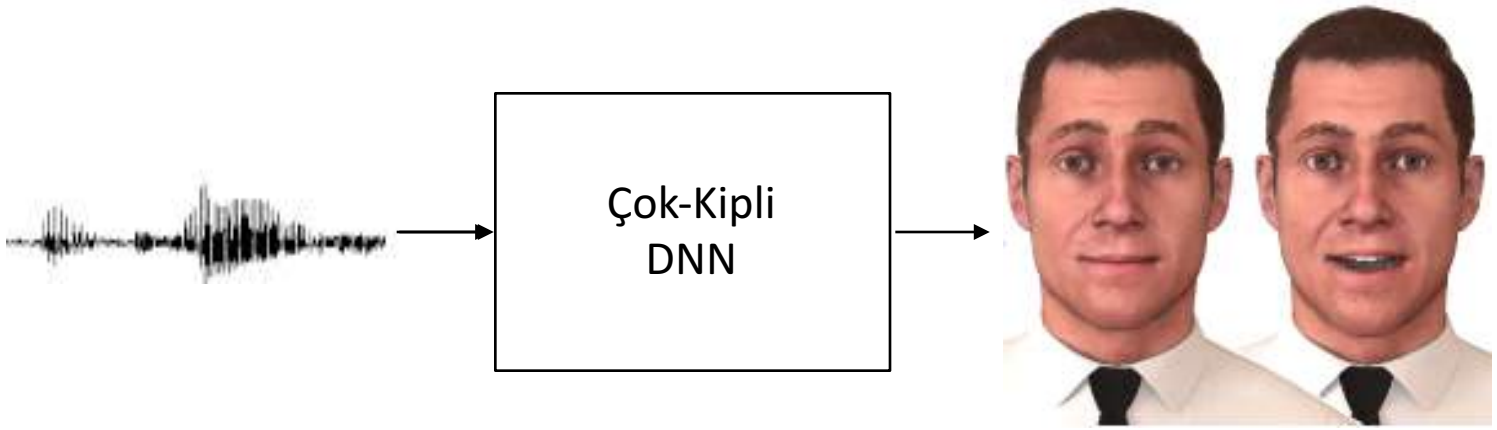
$$\{W_{T,k}, H_{T,k}\} = mGGYO(H_{T,k-1})$$

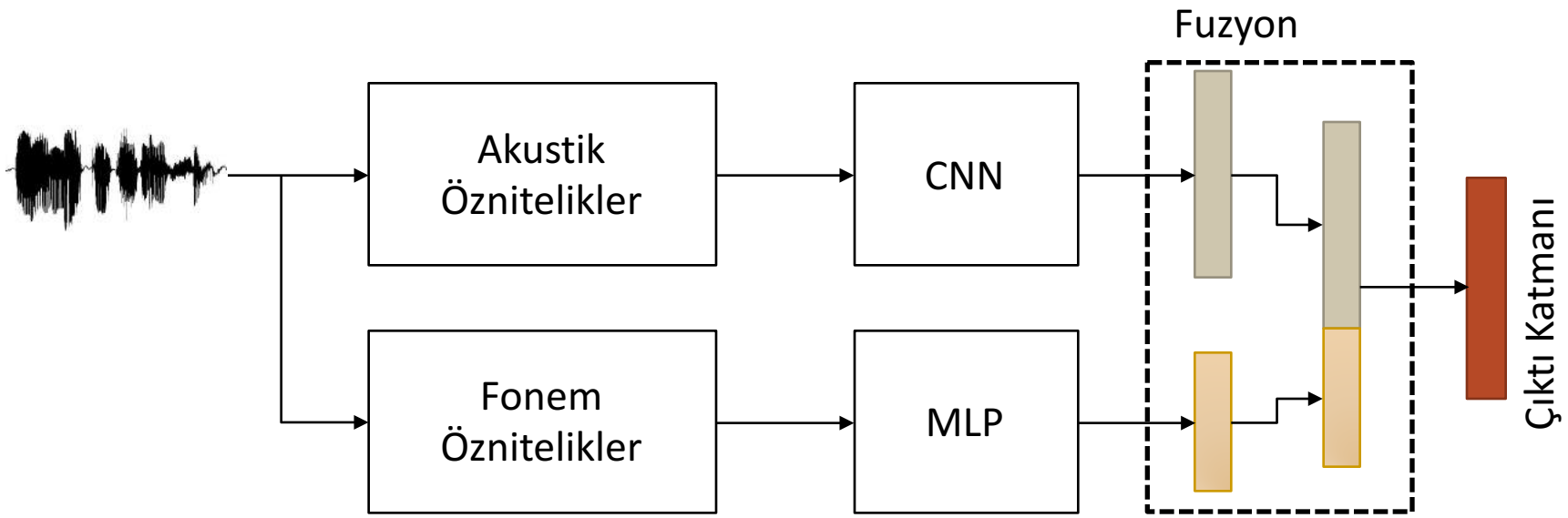
$$G_k = (H_{S,k}^c {H_{T,k}^c}^T)(H_{T,k}^c {H_{T,k}^c}^T + \lambda I)^{-1}$$

Konuşma-Yüz Modeli

- Problem: Duygudurumsal konuşma ile sürülen yüz sentezi
- Eldeki olası veriler
  - Kaynak / Duygudurum içermeyen veriler
    - GRID, BBC-LRW
  - Hedef / Duygudurum içeren veriler
    - MSP, SAVEE
  - Öznitelikler
    - Akustik öznitelikler
    - Dudak çevresi (alt-yüz) noktaları

Konuşma-Yüz Modeli Çok-Kipli Derin Yaklaşım (1)

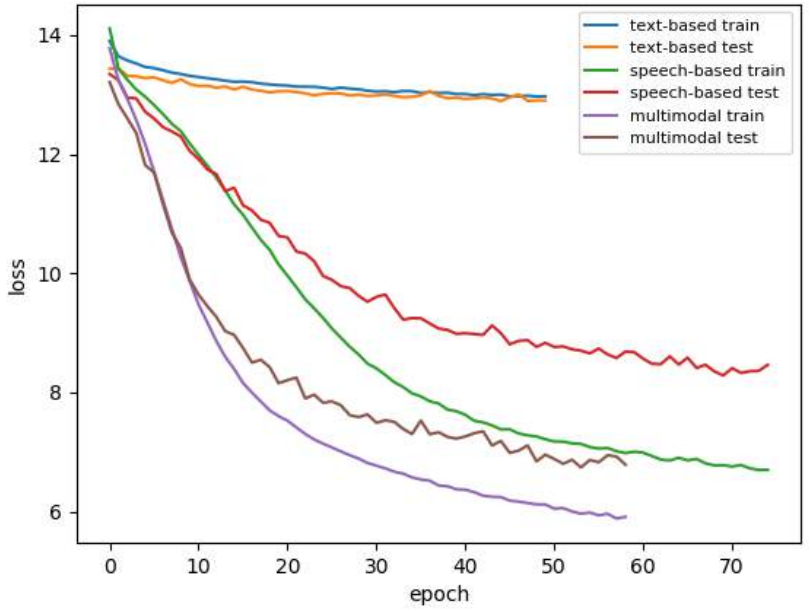


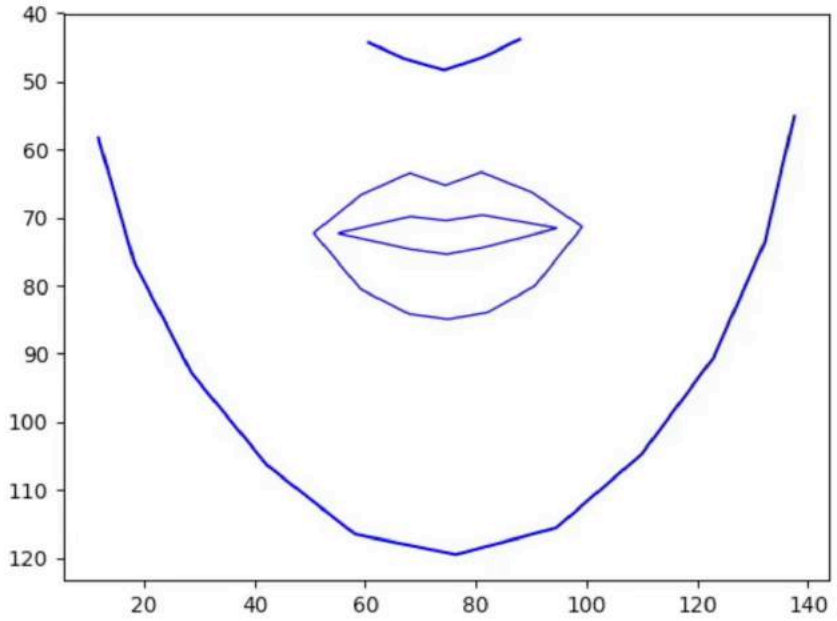


Konuşma-Yüz Modeli Çok-Kipli Derin Yaklaşım (2)

Erken Sonuçlar

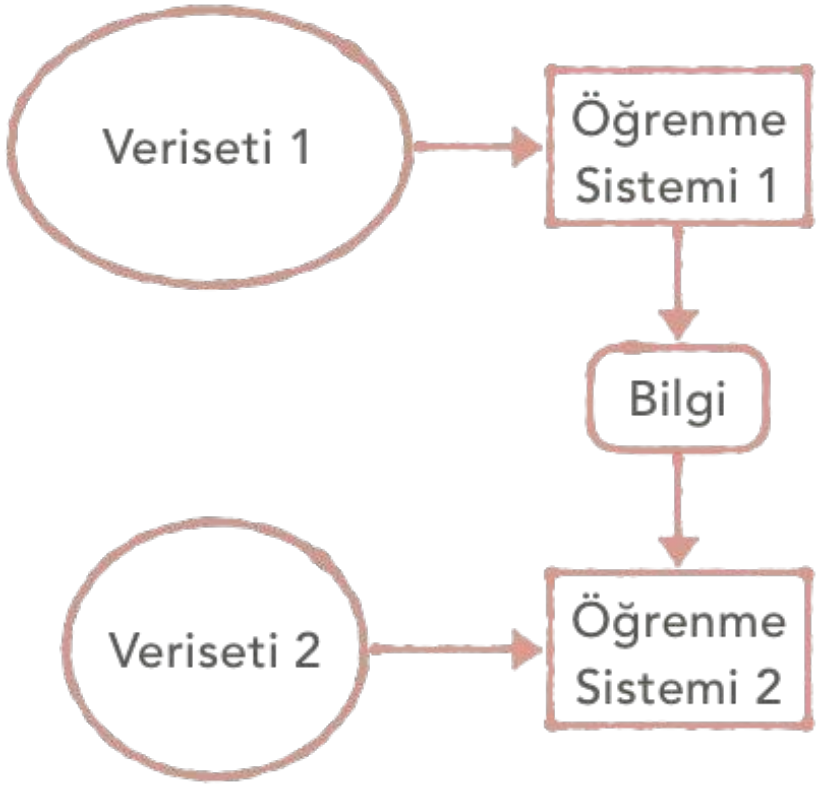
- SAVEE üzerinde çok-kipli kazanç
- GRID üzerinde metinden ve konuşmadan başarımlar benzer





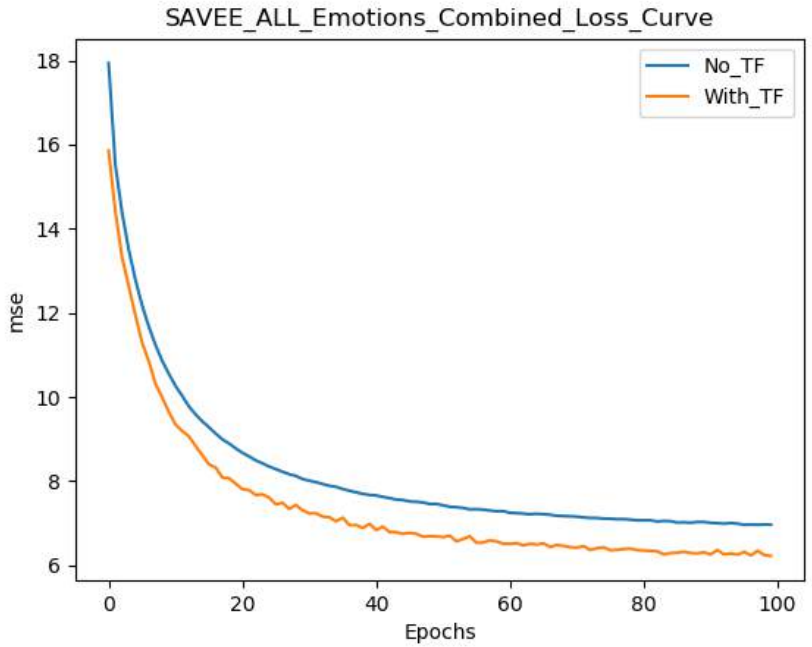
Konuşma-Yüz Modeli: Homojen Aktarmalı Öğrenme

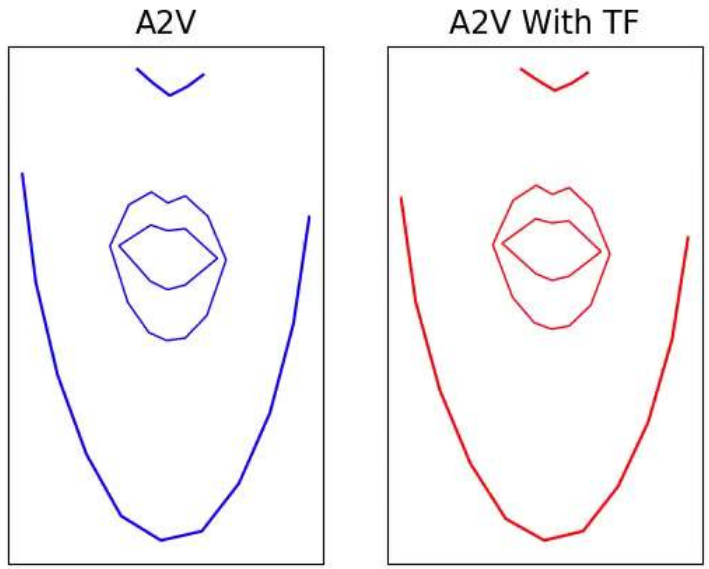
- Duygudurumsal konuşma ile sürülen yüz sentezi
  - Duygudurum içermeyen geniş veri ile ilklendir
  - Duygudurum içeren veriler ile uyarla



Konuşma-Yüz Modeli: Homojen Aktarmalı Öğrenme

Erken Sonuçlar

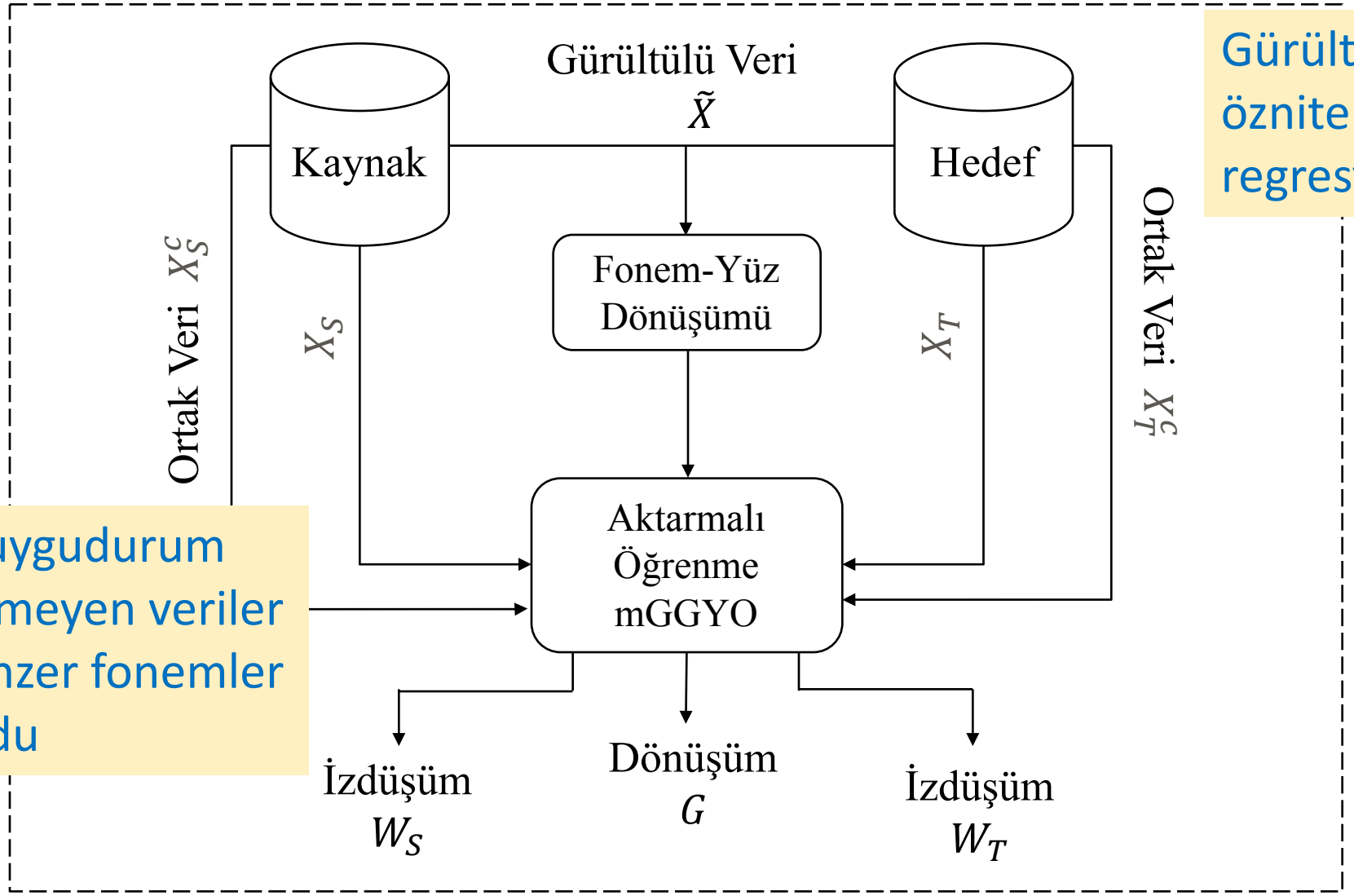




mGGYO Tabanlı Konuşma-Yüz Modeli (1)

- Problem: Duygudurumsal konuşma ile sürülen yüz sentezi
- Kaynak / Duygudurum içermeyen veriler
  - GRID, BBC-LRW
- Hedef / Duygudurum içeren veriler
  - MSP, SAVEE
- Öznitelikler
  - Akustik öznitelikler ve alt-yüz noktaları

mGGYO Tabanlı Konuşma-Yüz Modeli (2)



Gürültülü Veri: Metinden öznitelik vektörlerine regresyon ile oluşturuldu

Ortak Veri: Duygudurum içeren ve içermeyen veriler üzerinden benzer fonemler için oluşturuldu

mGGYO Tabanlı Dar-Bant Konuşma Tanıma (1)

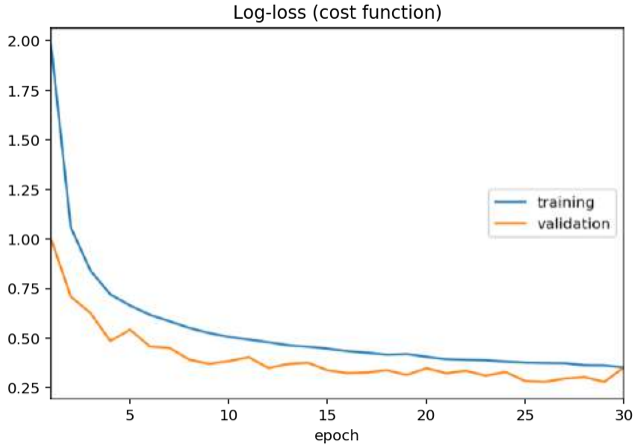
- **Problem**:
  - Bant genişliği düşük olan kayıtlar üzerinde konuşma tanıma
- **Motivasyon:**
  - Akustik olmayan gırtlak mikrofonları (GM) konuşmaları doku titreşimleri üzerinden yakalar
  - Ancak GM daha düşük frekanslarda dar bant genişliği sunar
- **Ön Çalışma**:
  - Bant genişliği GM seviyesine düşürülmüş sınırlı veri üzerinde mGGYO aktarmalı öğrenme sisteminin sınanması

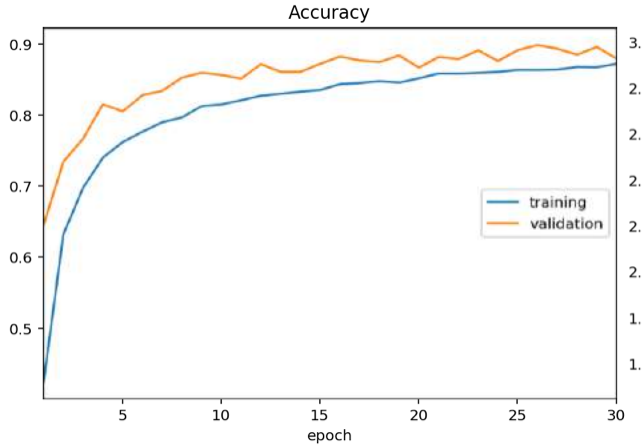
mGGYO Tabanlı Dar-Bant Konuşma Tanıma (2)

- Kaynak
  - TIMIT veritabanı (16kHz örnekleme)
- Hedef
  - GM bandında indirgenmiş TIMIT kayıtları (8kHz örnekleme)
- Görev
  - Fonem tanıma (39 sınıflı)
- Öznitelikler
  - Akustik mel-bant log-enerjiler

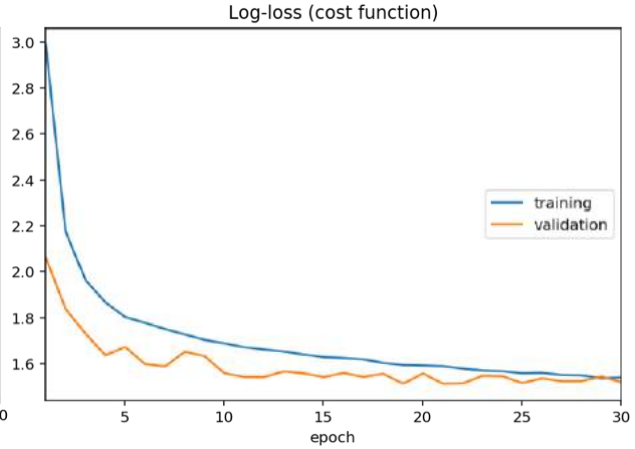
mGGYO Tabanlı Dar-Bant Konuşma Tanıma (3)

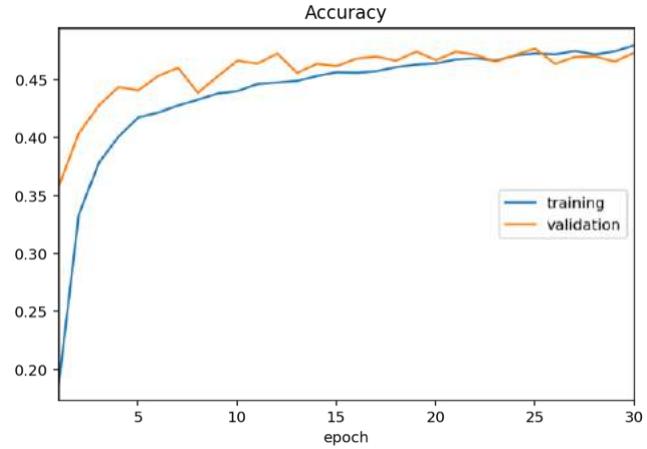
Geniş Bant



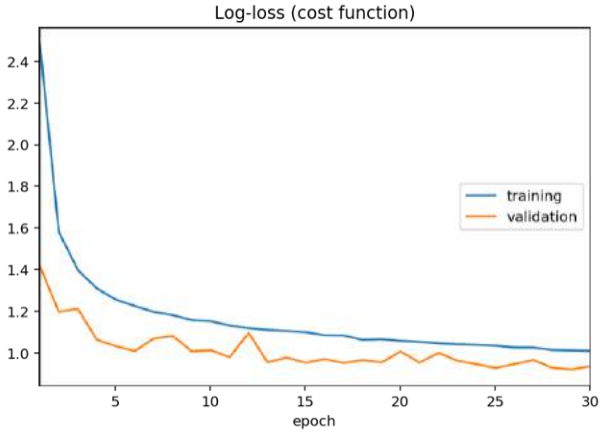


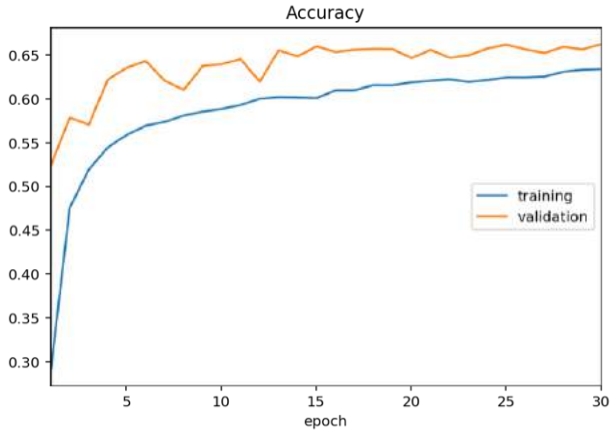
Dar Bant





Dar Bant - mGGYO

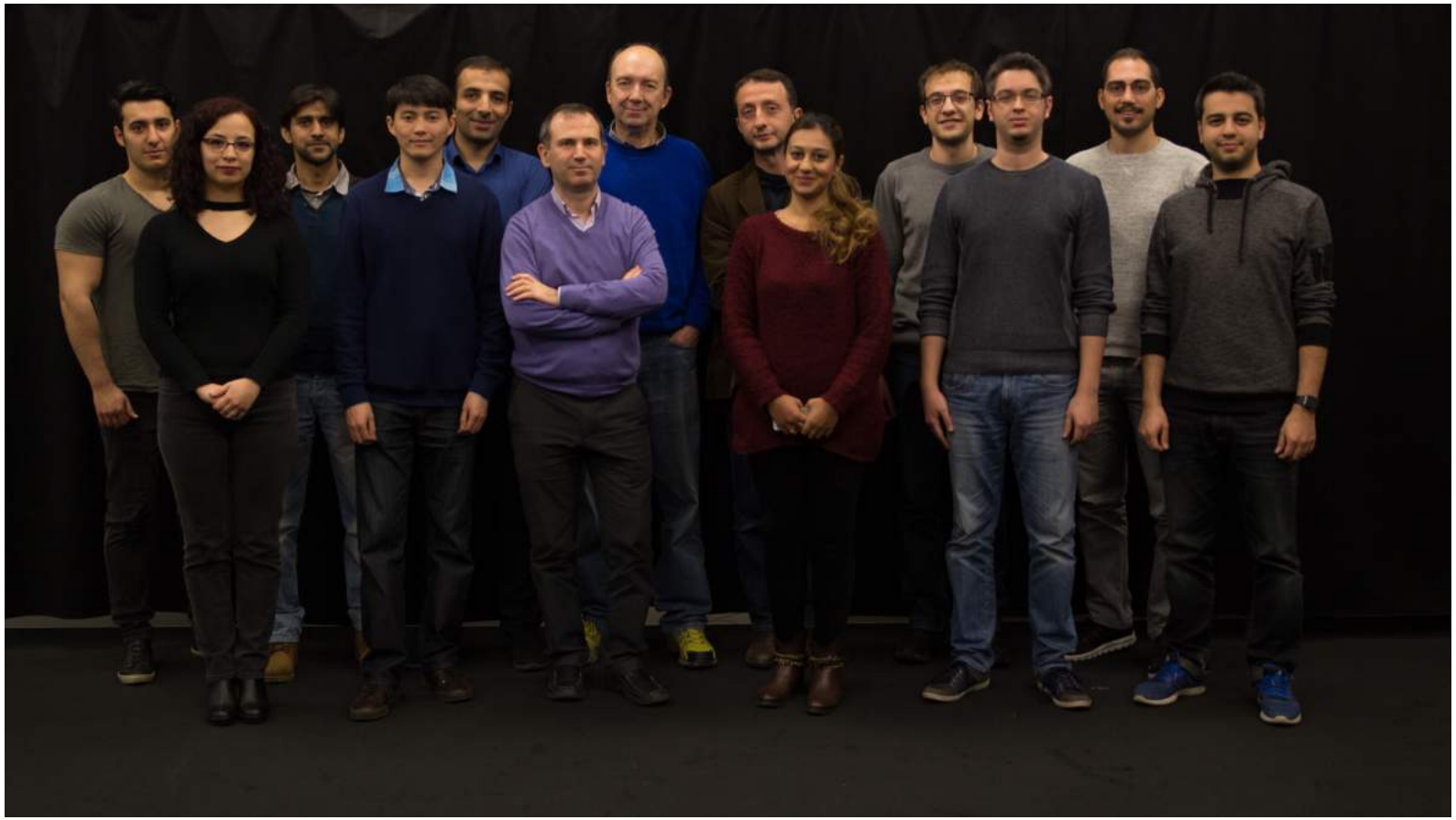




Vargılar

- Çok kipli konuşma işmar modellerine baktık
  - Farklı kipler farklı uygulamalar
    - Konuşma, el/kol/yüz/kafa/ses yolu hareketleri
  - Farklı ilinti yapıları için benzer ilinti modelleri
  - Benzer zorluklar
    - Veri azlığı
- Aktarmalı öğrenme
  - Duygulanımsal konuşmadan alt-yüz sentezi
  - Dar-bant konuşma tanıma

Teşekkürler



MVGL: Multimedya, Görü ve Grafik Laboratuvarı
mvgl.ku.edu.tr

Meraklısına

- J. T. Zhou, S. J. Pan, I. W. Tsang, Y. Yan. ”Hybrid Heterogeneous Transfer Learning through Deep Learning.” Proc. of the 28th conference on AI, 2014.
- K. Weiss, T. M. Khoshgoftaar, D. Wang. "A survey of transfer learning." Journal of Big Data, 2016.
- D. Wang, T. F. Zheng. "Transfer learning for speech and language processing." IEEE Signal and Information Processing Association Annual Summit and Conference, 2015.
- J. Yosinski, J. Clune, Y. Bengio, H. Lipson. "How transferable are features in deep neural networks?" Advances in Neural Information Processing Systems, 2014.
- X. Glorot, A. Bordes, Y. Bengio. "Domain adaptation for large-scale sentiment classification: A deep learning approach." International Conference on Machine Learning, 2011.
- S. J. Pan, Q. Yang. "A survey on transfer learning." IEEE Transactions on Knowledge and Data Engineering, 2010.
- M. Taylor, P. Stone. "Transfer learning for reinforcement learning domains: A survey." Journal of Machine Learning Research, 2009.
- Taylor, S. and et all. "A deep learning approach for generalized speech animation.” ACM Transactions on Graphics, 2017