

# Ultrason Görüntülerinden Parçacık Tabanlı Dil Konturu Tespiti

## Tongue Contour Extraction from Ultrasound Images Using Image Parts

Enes Aslan<sup>1,3</sup>  
<sup>1</sup>Kuveyt Türk Katılım Bankası,  
AR-GE Merkezi,  
Kocaeli, Türkiye  
enes.aslan@kuveytturk.com.tr

Naci Dumlu<sup>2</sup>  
<sup>2</sup>Pendik Devlet Hastanesi  
Radyoloji Bölümü  
İstanbul, Türkiye  
naci.dumlu@saglik.gov.tr

Yusuf Sinan Akgül<sup>3</sup>  
<sup>3</sup>GIT Vision Lab,  
<http://vision.gtu.edu.tr/>,  
Bilgisayar Mühendisliği Bölümü  
Gebze Teknik Üniversitesi  
Kocaeli, 41400, Türkiye  
agkul@gtu.edu.tr

**Özetçe**—Çalışmada Evrişimsel Sinir Ağları (ESA, CNN) derin öğrenme yöntemi kullanılarak otomatik bir şekilde dilin konturunu çıkaran bir çalışma önerilmektedir. Geliştirilen derin sinir ağı, ultrason görüntüsü parçacıkları ile eğitilmiştir. Eğitim aşamasında ultrason görüntüsü parçacıkları merkez noktasının dilin konturuna olan en yakın uzaklığı sınıf etiketi olarak kullanılmıştır. Görüntülerin video çerçevelerinden oluşmasının avantajı kullanılarak konturun sıralı video çerçeveleri arasındaki değişimi de eğitim girdisi olarak kullanılmıştır. Test aşamasında ise kayan pencere yöntemi kullanılarak elde edilen görüntü parçacıkları eğitim modeline girdi olarak verilip mesafeye bağlı sınıflar elde edilmiştir. Sınıfı sıfıra yakın olan parçacık dilin konturuna yakın olan konum anlamına gelmektedir. Parçacıkların merkez noktaları kullanılarak dilin konturu için aday noktalar oluşturulmuştur. Bu noktalara regresyon ile 3. dereceden polinom uydurularak dilin konturu elde edilmiştir. El ile işaretlenen kontur ile otomatik bulunan kontur karşılaştırılarak çalışma tamamlanmıştır. Kullanılan veri kümesi deney ortamında çekilmiş ultrason videolarından oluşmaktadır.

**Anahtar Kelimeler** — Ultrason, Derin Öğrenme, Dilin Konturunun Çıkartılması.

**Abstract**—In this work we propose automatic tongue contour extraction from ultrasound images based on Convolutional Neural Networks (CNN). The developed deep neural network was trained with ultrasound image parts. In training step the distances of the image patch centers from nearest tongue contour was used as the class label. By taking advantage of the fact that images consist of video frames, the time information was used as training input parameter. In the test phase, the image parts obtained by using the sliding window method are given as input to the training model and distance classes are obtained. A class that is close to zero means a position close to the tongue contour. Candidate points for the tongue contour were created using the centers points of the parts. The tongue contour is extracted by fitting a 3<sup>rd</sup> degree polynomial on center points of candidates. The work is completed by comparing the manually marked contour with the automatically found contour. The dataset consists of ultrasound videos taken in the experimental environment.

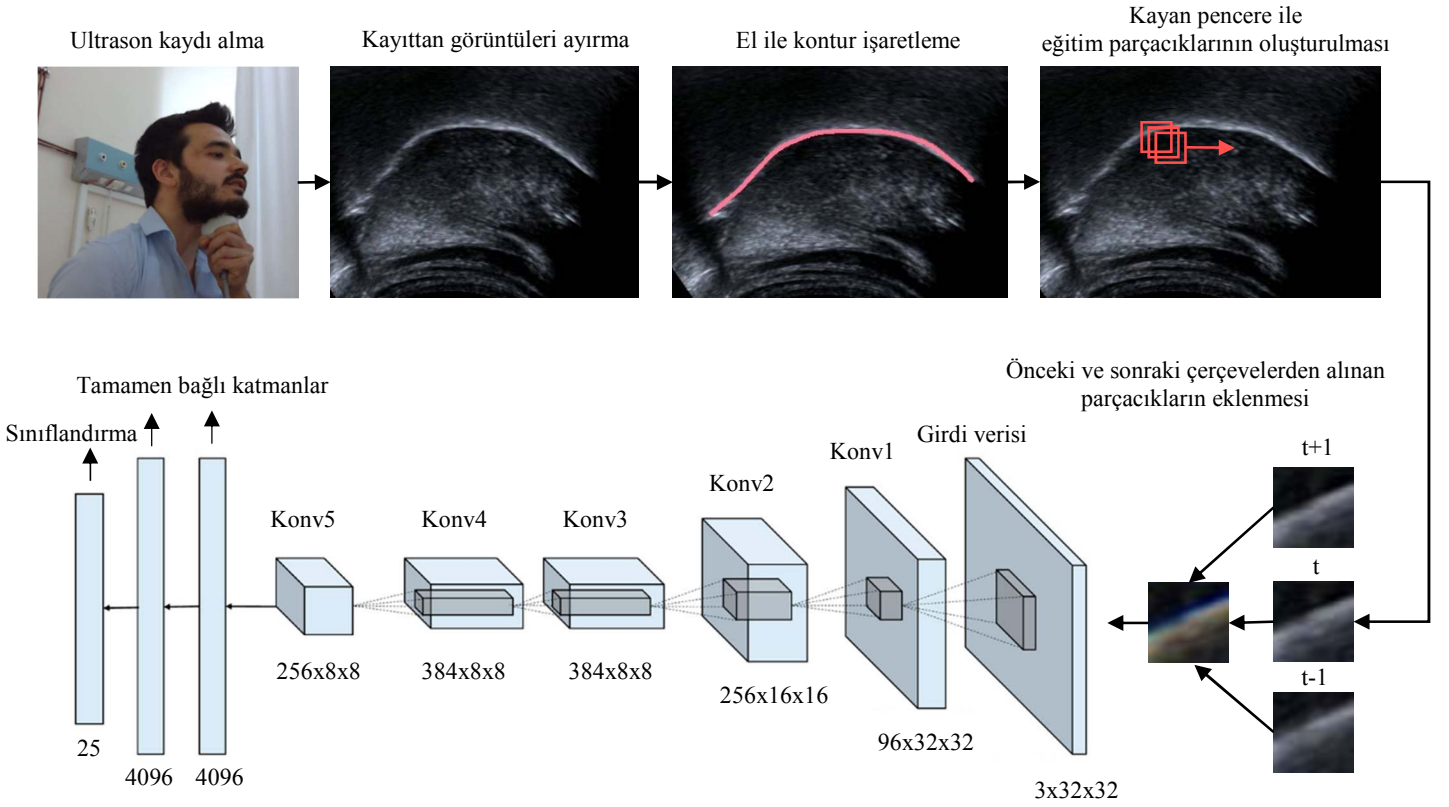
**Keywords** — Ultrasound, Deep Learning, Tongue Contour Extraction.

### I. GİRİŞ

Görüntülerden dilin konturu çıkartılarak; dilin modelinin oluşturulması, ses işleme, medikal ve biyometri gibi alanlarda kullanılmaktadır. Sesten yazıya dönüştürme, dilin yapısal bozukluğundan yola çıkarak bazı kalıtsal hastalıkların tespiti, konuşurken konturun değişiminden biyometrik veri çıkartılması bu alanlardaki uygulamalara örnek olarak verilebilir.

Öncelerde dilin konturunu bulma üzerine yapılan çalışmalarda [3], [4], [5] genelde aktif konturlar (Deformable Contours) [8] kullanılmıştır. Son zamanlarda yapılan çalışmalara bakıldığında ise derin öğrenme tabanlı yöntemlere [1], [2] rastlanmaktadır. Aktif konturlar ile yapılan çalışmalarda herhangi bir öğrenme olmaksızın resmin görüntü özellikleri kullanılarak konturun komşuluk ilişkileri, eğim, kenarlara yakınlık gibi parametrelerden oluşan bir enerji fonksiyonu en aza indirilmeye çalışılmaktadır. Yinelemelerle başta rastgele yerlerden başlatılan noktalar kümesi dilin konturuna yaklaşarak sabitlenmektedir. Derin öğrenme tabanlı yöntemlerden [2]'de SBM [7] (Sınırlı Boltzman Makineleri) derin sinir ağı ile iki fazlı otomatik kodlayıcı, otomatik kod çözücü mimarisi kullanılmıştır. Çalışma iki fazdan oluşmuştur. İlk fazda derin sinir ağı otomatik kodlayıcı olarak ultrason görüntüsünün tümü ve sadece kontur görüntüsü ile eğitilmiştir. İkinci fazda ise kod çözücü olarak ilk fazda eğitilmiş derin sinir ağına sadece tüm görüntü verilip kontur görüntüsü çıktı olarak alınmıştır. Bu çalışmada eğitim için verilen görüntüler el ile işaretlenmiştir. [1]'de ise aynı çalışmanın eğitim datasını otomatik işaretleyen bir sistem geliştirilmiştir. Blok eşleştirme adı verilen bir yöntem ile otomatik bir şekilde eğitim verisi işaretlenip ağ eğitilmiştir.

Şimdiye kadar yapılmış aktif konturlar tabanlı çalışmalar tam otomatik değildir, manuel ilklendirme ve çalışırken insan gözlemlemesi gerektirmektedir, yılan bir gürültüye takılırsa veya yolunu kaybederse insan müdahalesi ile yeniden ilklendirme yapılması gerekmektedir. Derin öğrenme tabanlı çalışmalarda ise görüntü parçacıkları işlenerek bir çözüm önerilmemiştir.



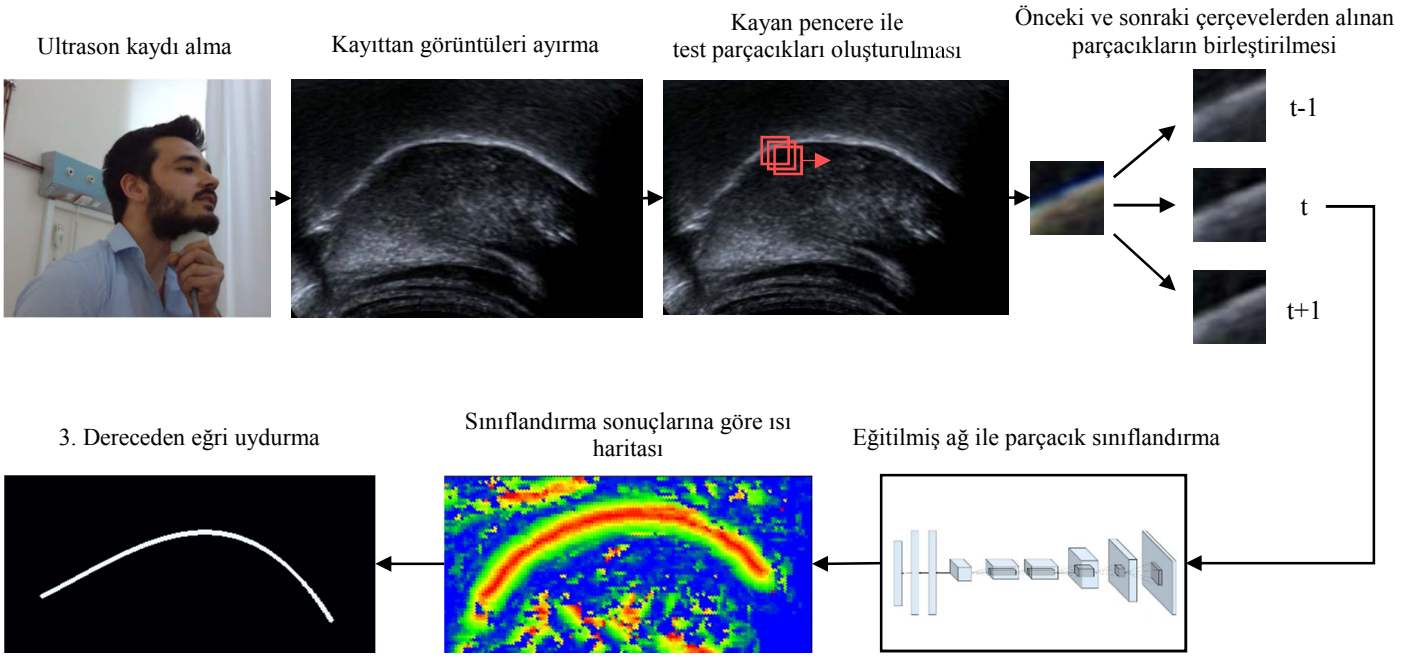
Şekil 1. Eğitim parçacıkları oluşturma ve eğitim.

Bu çalışmada ESA derin öğrenme yöntemi kullanılarak dilin ultrason görüntülerinden kontur elde edilmiştir. Son yıllarda nesne tanıma, görüntü bölütleme, ses tanıma, biyometri gibi alanlarda geleneksel makine öğrenmesi yöntemleri [6] yerine derin öğrenme tabanlı yöntemler popüler olmuştur. Derin öğrenme tabanlı yöntemler, büyük veri kümeleri üzerinde başarılı sonuçlar vermesinin yanında verilerden eğitim için özellikleri otomatik olarak çıkarmasıyla da bu alanda çalışan kişilere kolaylık sağlamaktadır. Bildiride önerilen çalışmada ESA, ultrason görüntüsü parçacıkları kullanılarak eğitilmiş, bu sayede eğitim kümesinden çok daha fazla özellik çıkartılarak gürbüz sonuçlar elde edilmiştir. Aynı zamanda veri kümesinin videolardan oluşmasının avantajı kullanılarak eğitime zaman bilgisi de eklenmiştir. Eğitilecek görüntü parçacığının merkez noktasının bir önceki ve bir sonraki çerçevede karşılık geldiği noktayı merkez kabul eden parçacıklar da derin sinir ağını eğitmede bir parametre olarak kullanılarak sonuçlar iyileştirilmiştir.

Bildirinin devamı şu şekildedir: 2. bölümde önerilen yöntem ile ilgili teknik detaylar ve uygulama hakkında bilgi verilmekte, 3. Bölümde deneylerde kullanılan veri kümesinin miktarı, test aşamasında kullanılan teknikler hakkında bilgi verilmekte, 4. bölümde ise çalışmanın sonuçları üzerine tartışma, değerlendirme ve gelecek çalışmalar hakkında bilgi verilmektedir.

## II. YÖNTEM

Önerilen sistem eğitim ve test olmak üzere iki aşamadan oluşmaktadır. İlk olarak deney ortamında çekilen ultrason videoları ayrıştırılarak görüntüler elde edilmiştir. Elde edilen bu görüntüler eğitim ve test kümesi olarak ikiye ayrılmıştır. Görüntülerin üzerindeki konturlar, geliştirilen bir işaretleme aracı ile eğitim ve test aşamalarında kullanılması için manuel olarak işaretlenmiştir. Bu sayede kontur üzerindeki her bir noktanın x, y pozisyonları elde edilmiştir. Bu pozisyonlar kullanılarak eğitim görüntüleri üzerinden rastgele alınan 32x32'lik parçacıklara, merkezinin kontura en yakın Öklid uzaklığına göre 1 ile 25 arasında sınıf atanmıştır. Buna göre Öklid uzaklığı 1 olan görüntü parçacıkları 1 sınıfına, 25 olan görüntü parçacıkları 25 sınıfına eklenmiştir. Eğitim aşamasında kullanılan ultrason görüntüsü çerçeveleri zamana göre sıralıdır. Kontura uzaklığına göre sınıf atanan bu görüntü parçacığının merkez pozisyonunun bir önceki ve bir sonraki çerçevelerde karşılığına denk gelen parçacıklar da alınıp bir eğitim verisi için Şekil 1'de görüldüğü gibi üç adet parçacık elde edilmiştir. Buna göre t, ana çerçeveyi, t-1 bir önceki çerçeveyi, t+1 ise bir sonraki çerçeveyi ifade etmektedir. Bu parçacıklar 32x32'lik üç kanallı boş bir görüntünün Kırmızı Yeşil Mavi (KYM)



Şekil 2. Sistemin test adımları.

kanallarına sırası ile eklenerek renkli, tek bir görüntü parçacığı elde edilmiştir. Her bir sınıf veri kümesi için bu işlem gerçekleştirilmiştir. Bu sayede zaman bilgisi de eğitime dahil edilmiştir.

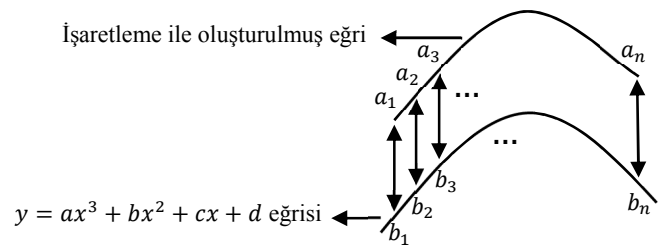
Sistemin eğitimi için Caffe [11] kütüphanesi Nvidia Digits arayüzü aracılığı ile kullanılmıştır. Yapay sinir ağı olarak ise AlexNet [10] kullanılmıştır. Ağ girdi olarak 256x256 boyutunda görüntüleri kabul etmektedir. Çalışmada kullanılan eğitim veri kümesi 32x32 boyutunda görüntülerden oluştuğu için özelleştirme gereksinimi duyulmuş ve girdi boyutu 32x32 olacak şekilde özelleştirilmiştir.

Eğitim aşamasında oluşturulan model kullanılarak Şekil 2’deki mimari ile testler gerçekleştirilmiştir. Test görüntüleri de eğitimde olduğu gibi zamana göre sıralıdır. Kayan pencere yöntemi ile görüntü üzerinden parçacıklar alınıp bir önceki ve bir sonraki görüntüde aynı pozisyona denk gelen parçacıklar, renkli ve boş bir görüntünün KYM kanallarına eklenerek tek bir parçacık oluşturulmuştur. Bu parçacık eğitilen modele verilerek 1-25 arası olan sınıfı belirlenmiştir. Görüntünün üzerindeki her bir nokta için bu işlem gerçekleştirilerek sınıfları bulunmuştur. İlk 1-7 sınıfa denk gelen parçacıkların merkez noktaları kullanılarak regresyon [9] ile 3.dereceden Denklem (1)’deki gibi bir polinom uydurulmuştur. Denklemde x ve y görüntü matrisinin satır ve sütunlarını, a, b, c, d ise sabit katsayıları ifade etmektedir.

$$y = ax^3 + bx^2 + cx + d \quad (1)$$

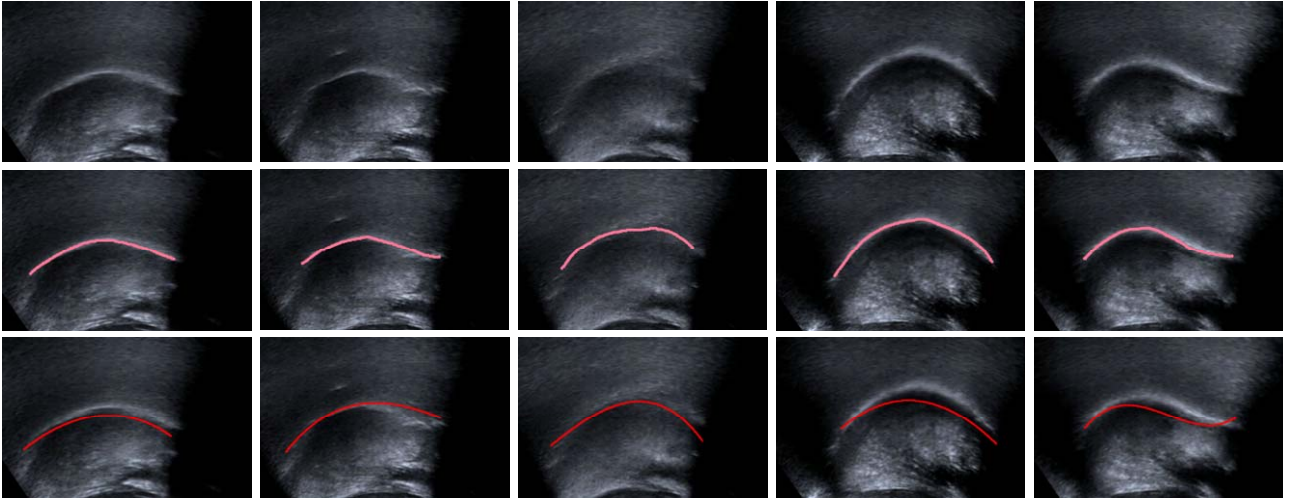
### III. DENEYLER

Deneyler, laboratuvar ortamında konuşma esnasında kaydedilmiş ultrason videolarından alınan görüntüler üzerinde gerçekleştirilmiştir. 4 denek ile 2 oturum gerçekleştirilmiştir. İlk oturumda alınan kayıtlar eğitim için, ikinci oturumda alınan kayıtlar ise test için kullanılmıştır. El ile işaretlenen 926 görüntüden 489’u eğitim 438’si test için kullanılmıştır. Eğitim için kullanılan 25 sınıfın veri kümesi, 489 görüntüden alınan rastgele parçacıklar ile her bir sınıf için 10.000 parçacık olmak üzere toplamda 25.000 parçacıktan oluşturulmuştur. Ağ eğitimi Nvidia’nın Tesla K20Xm grafik kartı kullanılarak gerçekleştirilmiştir. Caffe’ye Çözücü Tipi (Solver Type) olarak Olasılıksal Meyilli Azalım (Stochastic Gradient Descent) verilmiştir. Eğitim süresince veri kümesi ile 30 epok yapılmıştır.



Şekil 3. Hata hesabı.

$$\text{Ortalama Hata}(px) = \frac{\sum_{i=0}^n (a_i - b_i)}{n} \quad (2)$$



Şekil. 4. Farklı konturlar için örnek çıktılar, satırlar sırası ile ham görüntüler, el ile işaretlenmiş görüntüler ve sistemin bulduğu görüntülerdir.

Sistemin başarımının ölçümünde Denklem (2) kullanılmıştır. Buna göre;  $0 \leq i \leq n$  olmak üzere  $a_0$  noktası el ile işaretlenmiş konturun  $x$  eksenindeki başlangıç değeri ve buna karşılık gelen  $y$  değeri ikilisinden oluşmuş noktayı  $a_n$  ise aynı şekilde oluşturulmuş bitiş noktasını ifade etmektedir.  $b_0$ -  $b_n$  arasındaki noktalar ise sistemin bulduğu konturun üzerine uydurulmuş Denklem (1)'deki 3.dereceden polinomun  $x = 0$ ,  $x = n$  değerleri arası içindeki kalan noktalarını ifade etmektedir (Şekil 3). Her bir  $a, b$  noktaları arası Öklit uzaklık hesaplanarak toplanıp bir kontur için toplam nokta sayısına bölünerek hata hesabı yapılmıştır. Ultrason cihazından alınan veriler ile yapılan hesaplamalarda 1 pikselin yaklaşık olarak 0.241 milimetreye karşılık geldiği görülmüş ve ortalama hata milimetre türünden de hesaplanmıştır.

TABLO I. SONUÇLAR

İlk n Sınıf	Ortalama Hata (px)	Ortalama Hata (mm)
1	21.42	5.16
2	20.97	5.05
3	20.79	5.01
4	20.92	5.04
5	21.78	5.24
6	21.99	5.29
7	22.29	5.37

Deneyle tahminlerden elde edilen en iyi ilk 1-7 arası sınıf alınarak sıralı bir şekilde yapılmıştır. En düşük hata oranının ilk 3 sınıf alındığında elde edildiği gözlenmiştir. Deneyle kullanılan görüntüler 960x720 çözünürlüğündedir.

#### IV. SONUÇ

Bu çalışmada parçacık tabanlı ESA ile ultrason görüntülerinden dil konturu tespiti gerçekleştirilmiştir. Veri kümesinin videolardan oluşmasının avantajı kullanılarak eğitime zaman

bilgisi de eklenmiştir. Test aşamasında her piksel için işaretlenen nokta ile sistemin bulduğu nokta arasında hata hesabı yapıлып son aşamada ortalama hatayı hesaplayan bir ölçüm metriği kullanılmıştır. Geliştirilen sistemin tespit edilmesi zor, gürültülü konturları da başarı bir şekilde tespit ettiği gözlenmiştir. Sonraki çalışmalarda, eğitilen model zenginleştirilip birden fazla ölçek ile konturun çevresinden de bilgi alınarak başarı artırılabilir.

#### KAYNAKLAR

- [1] Jaumard-Hakoun, Aurore, et al. "Tongue contour extraction from ultrasound images based on deep neural network." *arXiv preprint arXiv:1605.05912* (2016).
- [2] Fasel, Ian, and Jeff Berry. "Deep belief networks for real-time extraction of tongue contours from ultrasound during speech." *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010.
- [3] Akgul, Yusuf Sinan, Chandra Kambhmettu, and Maureen Stone. "Automatic extraction and tracking of the tongue contours." *IEEE Transactions on Medical Imaging* 18.10 (1999): 1035-1045.
- [4] Li, Min, Chandra Kambhmettu, and Maureen Stone. "Automatic contour tracking in ultrasound images." *Clinical linguistics & phonetics* 19.6-7 (2005): 545-554.
- [5] Stone, Maureen. "A guide to analysing tongue motion from ultrasound images." *Clinical linguistics & phonetics* 19.6-7 (2005): 455-501.
- [6] Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up?: sentiment classification using machine learning techniques." *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*. Association for Computational Linguistics, 2002.
- [7] Hinton, Geoffrey E. "A practical guide to training restricted Boltzmann machines." *Neural networks: Tricks of the trade*. Springer, Berlin, Heidelberg, 2012. 599-619.
- [8] Lai, Kok F., and Roland T. Chin. "Deformable contours: Modeling and extraction." *IEEE Transactions on pattern analysis and machine intelligence* 17.11 (1995): 1084-1090.
- [9] Motulsky, Harvey, and Arthur Christopoulos. *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*. Oxford University Press, 2004.
- [10] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.
- [11] Jia, Yangqing, et al. "Caffe: Convolutional architecture for fast feature embedding." *Proceedings of the 22nd ACM international conference on Multimedia*. ACM, 2014.