An Improved Air Tissue Boundary Segmentation Technique for Real Time Magnetic Resonance Imaging Video Using SegNet

Valliappan CA[1], Avinash Kumar[2], Renuka Mannem[1], Karthik GR[1], Prasanta Kumar Ghosh[1]

[1]SPIRE LAB, Electrical Engineering
Indian Institute of Science (IISc), Bangalore, Karnataka, India
[2]Electrical and Electronics Engineering,
National Institute of Technology, Surathkal, Karnataka, India

SPIRE LAB

May 17, 2019

Section 1

Introduction

- **Goal**: Segmentation of the Air-Tissue Boundaries (ATBs) with minimum number of training videos

Introduction

- **Goal**: Segmentation of the Air-Tissue Boundaries (ATBs) with minimum number of training videos

Introduction

- **Goal**: Segmentation of the Air-Tissue Boundaries (ATBs) with minimum number of training videos

- **Approach**: Semantic segmentation using Segmentation Network (SegNet).

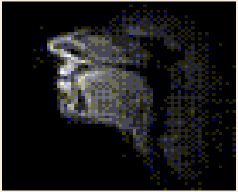
Motivation

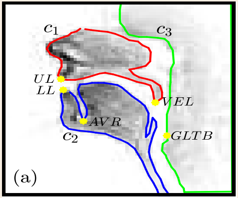
Need for study

Understanding speech production.

Dataset

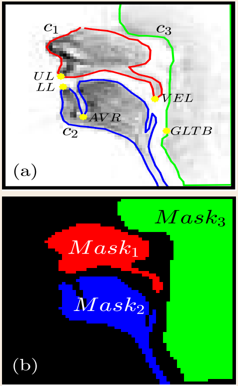
- **USC-TIMIT** corpus
- **2-Female** (F1, F2) and **2-Male** (M1, M2).
- Subset : 16 Videos from each subject.
- Video : 23.18 fps
- Spacial resolution of 68 68.





Dataset

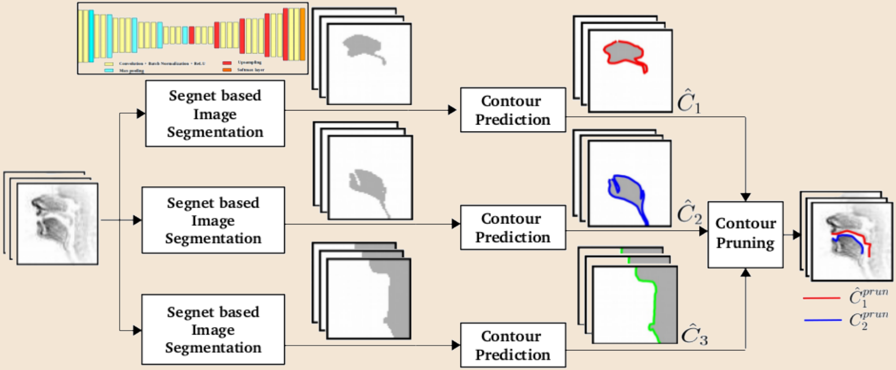
- **Manual annotation:**
  1. Complete ATBs
  2. Upper lip (UL)
  3. Lower lip (LL)
  4. Tongue base (AVR)
  5. Velum tip (VEL)
  6. Glottis begin (GLTB)
- Number of frames: 1462, 1270, 1642, 1399 for subjects F_1, F_2, M_1, M_2 respectively.
- Division of tissue regions into **3 masks**.



Section 2

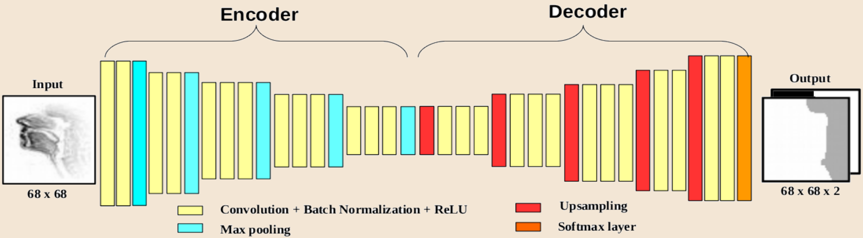
Proposed SegNet based Approach

Illustration of the steps in the proposed SegNet based approach



Proposed SegNet based segmentation

SegNet architecture[1]

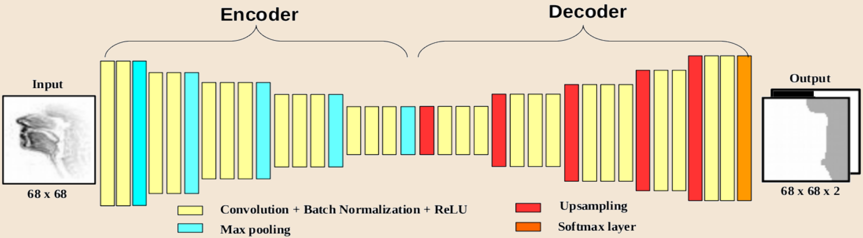


1. Symmetric encoder-decoder.

[1] Karen et. al, "Very Deep Convolutional Networks for Large-Scale Image Recognition," CoRR, 2014.

Proposed SegNet based segmentation

SegNet architecture[1]

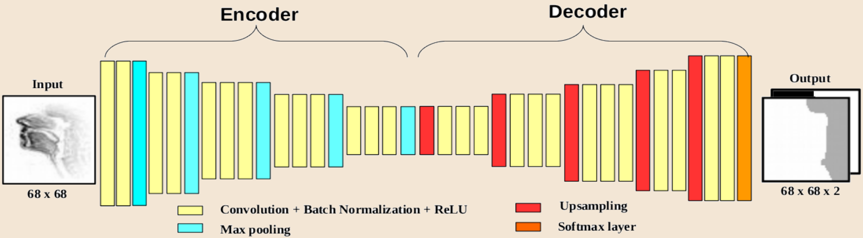


1. Symmetric encoder-decoder.
2. **Three Segnets:** One SegNet for each mask.

[1] Karen et. al, "Very Deep Convolutional Networks for Large-Scale Image Recognition," CoRR, 2014.

Proposed SegNet based segmentation

SegNet architecture[1]



1. Symmetric encoder-decoder.
2. **Three Segnets:** One SegNet for each mask.
3. $SegNet_i$: Does a given pixel belong to $mask_i$ or air cavity region?

[1] Karen et. al, "Very Deep Convolutional Networks for Large-Scale Image Recognition," CoRR, 2014.

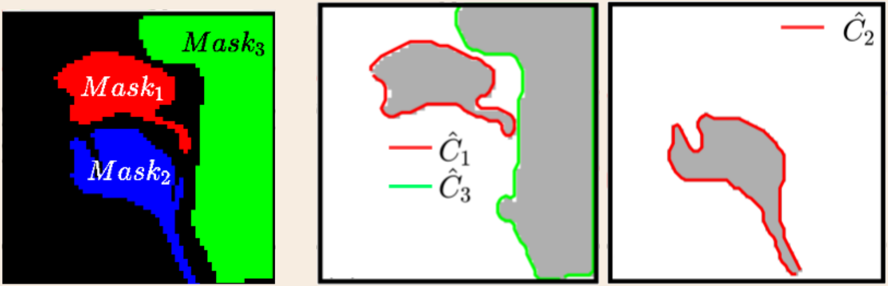
Contour Prediction

- **Stage 1:** Canny edge detection
- **Stage 2:** Connecting edges via concave hull algorithm [1]

[1] J.-S. Park et. al "A new concave hull algorithm and concaveness measure for n-dimensional datasets", 2018

Contour Prediction

- **Stage 1:** Canny edge detection
- **Stage 2:** Connecting edges via concave hull algorithm [1]



[1] J.-S. Park et. al "A new concave hull algorithm and concaveness measure for n-dimensional datasets", 2018

Contour Pruning

Contour Pruning

Obtain upper and lower contours within the vocal tract

Contour Pruning

Obtain upper contour within vocal tract:

Contour Pruning

Obtain lower contour within vocal tract:

Contour Pruning

Obtain lower contour within vocal tract:

Contour Pruning

Obtain lower contour within vocal tract:

Contour Pruning

Obtain lower contour within vocal tract:

2^{nd} order polynomial fit

Proposed SegNet based Approach

Illustration of the steps in the proposed SegNet based approach

Section 3

Experimental Setup

Baselines:

- Maeda grid-line[1] (MG).

[1] Kim et.al, "Enhanced airway-tissue boundary segmentation for real-time magnetic resonance imaging data," ISSP, 2014.

Experimental Setup

Baselines:

- Maeda grid-line[1] (MG).
- Fisher-discrimination measure based segmentation[2] (FDM)

[1] Kim et.al, "Enhanced airway-tissue boundary segmentation for real-time magnetic resonance imaging data," ISSP, 2014.

[2] A. Koparkar et. al, "A supervised air-tissue boundary segmentation technique in real-time magnetic resonance imaging video using a novel measure of contrast and dynamic programming," ICASSP, 2018.

Experimental Setup

Baselines:

- Maeda grid-line[1] (MG).
- Fisher-discrimination measure based segmentation[2] (FDM)
- fully convolutional networks based segmentation[3] (FCN)

[1] Kim et.al, "Enhanced airway-tissue boundary segmentation for real-time magnetic resonance imaging data," ISSP, 2014.

[2] A. Koparkar et. al, "A supervised air-tissue boundary segmentation technique in real-time magnetic resonance imaging video using a novel measure of contrast and dynamic programming," ICASSP, 2018.

[3] Valliappan CA et. al, Air-tissue boundary segmentation in real-time magnetic resonance imaging video using semantic segmentation with fully convolutional networks," Interspeech, 2018

Experimental Setup-1 for ATB Estimation

- 4-fold setup

Experimental Setup-1 for ATB Estimation

- 4-fold setup
- Training set : 2900
- Development & Test set : 1443

Experimental Setup-1 for ATB Estimation

- 4-fold setup
- Training set : 2900
- Development & Test set : 1443
- 30 epochs, early stopping condition.

Experimental Setup-1 for ATB Estimation

- 4-fold setup
- Training set : 2900
- Development & Test set : 1443
- 30 epochs, early stopping condition.

Experimental Setup-2

- Estimating the minimum number of rtMRI videos required for training for FCN and SegNet.

Experimental Setup-2

- Estimating the minimum number of rtMRI videos required for training for FCN and SegNet.
- 8 Models of FCN and SegNet
- The i^{th} model - i training videos from four subjects, where $i \in \{1, 2, \ldots, 8\}$.

Experimental Setup-2

- Estimating the minimum number of rtMRI videos required for training for FCN and SegNet.
- 8 Models of FCN and SegNet
- The i^{th} model - i training videos from four subjects, where $i \in \{1, 2, \ldots, 8\}$.
- Each video - 90 frames
- Fixed Development & Test set : 1443
- 30 epochs, early stopping condition.

Evaluation metrics

- DTW distance [1]: Measures the closeness of the estimated contour the ground truth contour (unit:pixel).

[1] Berndt et. al, "Using dynamic time warping to nd patterns in time series," KDD, 1994.

Evaluation metrics

- DTW distance [1]: Measures the closeness of the estimated contour the ground truth contour (unit:pixel).

- Pixel accuracy[2]: To evaluate the performance of FCN and SegNet.

[1] Berndt et. al, "Using dynamic time warping to nd patterns in time series," KDD, 1994.

[2] J. Long et. al, "Fully convolutional networks for semantic segmentation", 2015.

Section 4

DTW distances (Upper ATB)

	Upper ATB							
SUB	MG		FCN		SegNet		FDM	
F_1	1:02	0:19	0:91	0:21	0.83	0.11	0:94	0:17
F_2	1:24	0:29	1:08	0:19	0.96	0.15	1:16	0:19
M_1	1:10	0:20	1.02	0.20	1:15	0:16	1:11	0:20
M_2	1:19	0:24	1.09	0.21	1:10	0:19	1:10	0:23
AVG:	1:13	0:22	1:02	0:20	1.02	0.15	1:08	0:19

Average (standard deviation) DTW distance of the predicted upper ATBs within the vocal tract

DTW distances (Upper ATB)

	Upper ATB							
SUB	MG		FCN		SegNet		FDM	
F_1	1:02	0:19	0:91	0:21	0.83	0.11	0:94	0:17
F_2	1:24	0:29	1:08	0:19	0.96	0.15	1:16	0:19
M_1	1:10	0:20	1.02	0.20	1:15	0:16	1:11	0:20
M_2	1:19	0:24	1.09	0.21	1:10	0:19	1:10	0:23
AVG:	1:13	0:22	1:02	0:20	1.02	0.15	1:08	0:19

Average (standard deviation) DTW distance of the predicted upper ATBs within the vocal tract

- SegNet yields better or comparable performance relative to baselines.

DTW distances (Lower ATB)

	Lower ATB							
SUB	MG		FCN		SegNet		FDM	
F_1	1:21	0:21	1:00	0:25	0.92	0.17	0:99	0:23
F_2	1:28	0:27	1:13	0:31	1.12	0.29	1:24	0:25
M_1	1:26	0:60	1:17	0:25	1.16	0.26	1:17	0:26
M_2	1:35	0:30	1:21	0:23	1:18	0:24	1.16	0.41
AVG:	1:27	0:35	1:13	0:26	1.09	0.23	1:14	0:29

Average (standard deviation) DTW distance of the predicted lower ATBs within the vocal tract

DTW distances (Lower ATB)

	Lower ATB							
SUB	MG		FCN		SegNet		FDM	
F_1	1:21	0:21	1:00	0:25	0.92	0.17	0:99	0:23
F_2	1:28	0:27	1:13	0:31	1.12	0.29	1:24	0:25
M_1	1:26	0:60	1:17	0:25	1.16	0.26	1:17	0:26
M_2	1:35	0:30	1:21	0:23	1:18	0:24	1.16	0.41
AVG:	1:27	0:35	1:13	0:26	1.09	0.23	1:14	0:29

Average (standard deviation) DTW distance of the predicted lower ATBs within the vocal tract

- SegNet yields better or comparable performance relative to baselines.

Complete ATBs

	C_1		C_2		C_3	
SUB	SegNet	FCN	SegNet	FCN	SegNet	FCN
F_1	0.88	0:89	0.85	1:05	0.80	0:83
F_2	0.98	1:02	1:15	1.12	0:81	0.80
M_1	1:03	1:03	0.94	1:37	0.79	0:80
M_2	1:03	0.89	1:03	1.01	0.83	0:85

Average DTW distance of the predicted complete ATBs for all the subjects

Pixel Accuracy For the SegNet and FCN models

SUB	$Model_1$	$Model_2$	$Model_3$	$Model_4$	$Model_5$	$Model_6$	$Model_7$	$Model_8$
$Mask_1^{seg}$	88.70	99.54	99.53	99.57	99.54	99.54	99.55	99.57
$Mask_2^{seg}$	85.89	98.64	98.65	98.61	98.65	98.60	98.64	98.68
$Mask_3^{seg}$	90.30	99.78	99.77	99.77	99.76	99.76	99.78	99.77
$Mask_1^{fcn}$	85.68	90.89	94.47	96.09	98.14	99.17	99.24	99.28
$Mask_2^{fcn}$	84.12	88.14	93.88	95.51	97.77	98.09	98.08	98.14
$Mask_3^{fcn}$	89.45	93.45	95.80	98.80	99.60	99.71	99.73	99.72

Pixel classification accuracy averaged across all subjects (on test set) for each mask vs number of training videos for SegNet, FCN.(Bold indicating the saturation point)

Section 5

Discussion

1. On an average 0:70% pixels are being misclassi ed (unlike 1% for FCN).
2. Misclassi ed pixels {boundary region : due to low resolution of the image.
3. Precision of annotation : 1decimal place
4. Proposed method :Pixel level

Section 6

Conclusions

- Proposed method yields better performance than the baseline { DTW
  distance
- SegNet requires only two training videos per subject.

Conclusions

- Proposed method yields better performance than the baseline { DTW distance
- SegNet requires only two training videos per subject.

Future Directions

Data augmentation to further reduce the minimum number of training videos required for better pixel accuracy.

Section 7

The authors thank Pratiksha Trust for their support

Questions?