# Head and Neck Auto Segmentation Challenge based on Non-Local Generative Models

M. Orbes-Arteaga, D. Cárdenas-Peña, and G. Castellanos-Dominguez[1]

Signal Processing and Recognition Group - Universidad Nacional de Colombia
{hmorbesa,dcardenasp,cgcastellanosd}@unal.edu.co

## 1 Introduction

Multi-atlas segmentation approaches register a set of labeled atlases to a given target image and combine their contributions using a label fusion strategy to provide a labeled image. Although this kind of approaches attains suitable results on brain tissue classification tasks, registration issues can bias the label estimation. A non-local strategy -known as label fusion- reduces the misalignment influence by allowing the spatial neighbors of a voxel to vote for its label according to a weighting function. Therefore, the selection of such a function is a critical factor for achieving an accurate segmentation; such function can be either similarity-based [1] or reconstruction-based [2]. In general, both of above approaches demand a voxel representation, being the patch one the most commonly considered. However, those label fusion methodologies present the following issues: i) patch similarity and label affinity among voxels may be unrelated; ii) similarity measures are only based on intensity obviating the label information; and iii) although the atlas patches are fully labeled, only the central voxel of the target patch is classified.

We propose a new patch-based segmentation dealing with above issues. To this end, we weight the label votes using a generative probabilistic approach: The probability of the target patch being generated by the model of an atlas patch is computed; a Gibbs distribution spatially constrains the model parameters. This approach allows us to estimate the label of the whole target patch using the class conditional probability for each voxel. Finally, the use of overlapping neighborhoods leads to estimate several times the voxel labels. Hence, the multiple estimations are combined using a 3D sliding Gaussian window.

## 2 Method

Given a target image $\boldsymbol{X}^q$, the segmentation process consists in finding the label map $\boldsymbol{L}^q$ for the target image using a registered training dataset, $\mathcal{X}=\{\boldsymbol{X}^n,\boldsymbol{L}^n{:}n=1,\ldots,N\}$. The intensity and label images of the $n$-th atlas compose the pair $\{\boldsymbol{X}^n,\boldsymbol{L}^n\}$, where $\boldsymbol{X}^n=\{x_r^n\in\mathbb{R}{:}r\in\Omega\}$ and $\boldsymbol{L}^n=\{l_r^n\in[1,C]{:}r\in\Omega\}$, the value $r$ indexes the spatial elements, and $C$ is the number of possible classes.

For each voxel $r$, we extract a set of patches $\mathcal{P}_r=\{\beta_y^n\subset\boldsymbol{X}^n,\gamma_y^n\subset\boldsymbol{L}^n{:}y\in\eta(r)\}$, where $\eta(r)$ is a neighborhood around $r$, and denote the target patch as $\beta_r^q\subset\boldsymbol{X}^q$. In this sense, each patch is an arrange $\beta_y^n=\{x_s^n:\|y-s\|<\xi\}$, being $\xi$ the patch radius. In

order to reduce the computational time and remove outlying patches, we pre-select the most similar candidate patches by computing the structural similarity measurement *ss* [3]:

$$ss(y,r) = \left(\frac{2\mu_y\mu_r}{\mu_y^2+\mu_r^2}\right)\left(\frac{2\sigma_y\sigma_r}{\sigma_y^2+\sigma_r^2}\right) \tag{1}$$

$$\mathcal{D}_r = \left\{\beta_y^n : ss(y,r) \geq 0.9\right\} \tag{2}$$

Figure 1 shows the process for extracting patches.

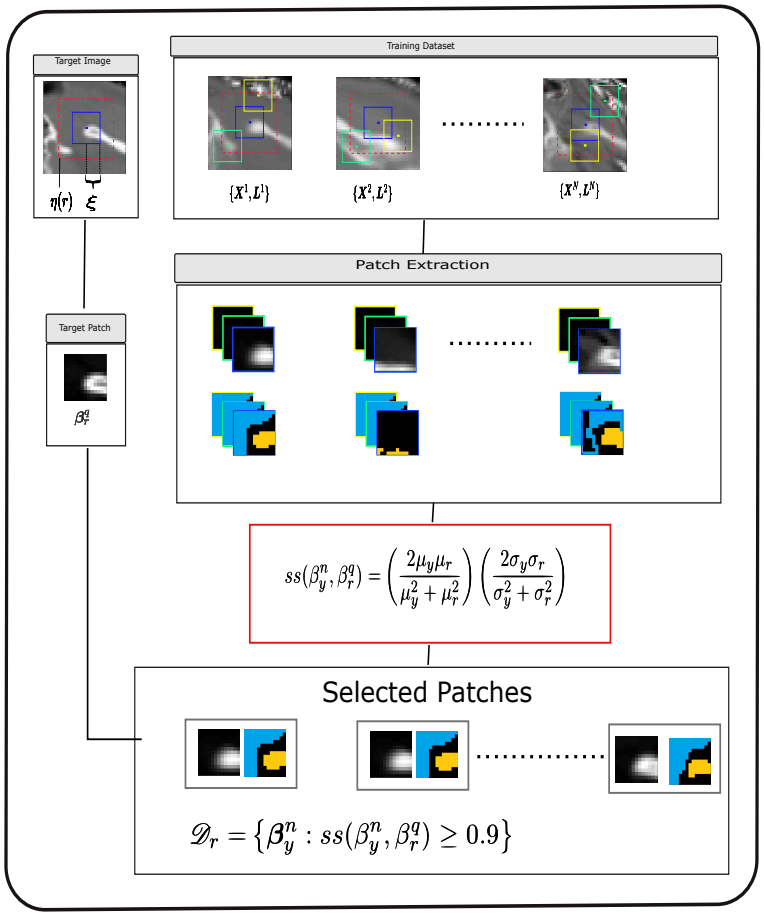


Fig. 1: Patches are extracted from the intensity and labeled images on the training dataset, structural similarity is used to select de most similar patches

### 2.1 Weighted computation

Aiming to take into account the information provided by the labels in the patch fusion step, we use a generative probabilistic method as a weighting criterion. The procedure is as follow. We propose to estimate the label $l_r^q$ of the $r$-th voxel as double weighted voting: the first one accounts for the contribution of the neighboring patches, the second one weights the overlapping neighborhoods. Equation (3) introduces the resulting label fusion,

$$l_r = \arg\max_c \sum_{r' \in \beta_r^n} k(r,r') \sum_{\beta_y^n \in \mathcal{D}_r} w_{ry}^n P(x_r|\Phi_{y,c}^n), \tag{3}$$

where $k(r,r') = \frac{1}{Z_k} \exp\left(\|r-r'\|/2h^2\right)$ is the sliding window with size $h \in \mathbb{R}^+$ and normalization constant $z_k \in \mathbb{R}^+$, $w_{ry}^n \in [0,1]$ is the similarity between $\beta_y^n$ and $\beta_r^q$, the probability of the $r$-th voxel to belong to the class $c$ and being generated by the patch $\beta_y^n$ is $P(x_r|\Phi_{y,c}^n) \in [0,1]$, and $\Phi_{y,c}^n$ are the parameters of the generative model for $\beta_y^n$. Particularly, we assume that each atlas patch is modeled by a mixture of Gaussian. Hence, $\beta_y^n$ is represented by the set of parameters $\Phi_y^n = \{\mu_{yc}^n, \sigma_{yc}^n, v_{yc}^n\}$ computed as:

$$\mu_{yc}^n = \boldsymbol{E}\{x_r^n | l_r^n = c : \|y-r\| < \xi\} \tag{4}$$

$$\sigma_{yc}^n = \boldsymbol{E}\{(x_r^n - \mu_{yc}^n)(x_r^n - \mu_{yc}^n)^T | l_r^n = c : \|y-r'\| < \xi\} \tag{5}$$

$$v_{yc}^n = \left\{ v_{rc}^n = \sum_{s \in \varepsilon(r)} 1 - \delta(l_s^n - c) : \|y-r'\| < \xi \right\}, \tag{6}$$

Where the 6-neighboring cliques, $\varepsilon(r)$, provide a spatial smoothness and $\delta(\cdot)$ is Dirac function. As a result, the probability $P(x_r|\Phi_{y,c}^n)$ is estimated as:

$$P(x_r|\Phi_{yc}^n) = \mathcal{N}(x_r|\mu_{yc}^n, \sigma_{yc}^n)\mathcal{G}(v_{rc}^n) \tag{7}$$

$\mathcal{N}(x|\mu,\sigma)$ is a Gaussian distribution with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma \in \mathbb{R}+$, and $\mathcal{G}(z) = \frac{1}{Z_G}\exp(-z)$ is the Gibbs distribution with normalization constant $Z_G \in \mathbb{R}^+$. Finally, assuming that the more similar the patches, the larger the probabilities, we define the patch-wise similarity $w_{ry}^n$ as the probability of the target patch being generated by the mixture of gaussians with parameters $\Phi_y^n$:

$$w_{ry}^n(\beta_r^q, \beta_y^n) = \prod_{s \in \beta_r^q} \sum_c \mathcal{N}(s|\mu_{y,c}^n, \sigma_{y,c}^n) P(v_{r,c}^n) \tag{8}$$

Figure 2 shows the complete segmentation process

## 3 Experiments And Results

### 3.1 Data

Our proposed method was evaluated on the Head and neck Auto-Segmentation challenge (MICCAI 2015). Three datasets which correspond to: training (25 CT images), off-site(10 CT images), and on-site(5 CT images) were provided for the challenge organizers. Training images contain the manual annotations of seven organs, namely Brainstem, Chiasm, Mandible, Optic Nerves, Parotid and Submandibular gland. Whereas, off-site and on-site datasets were used for testing.

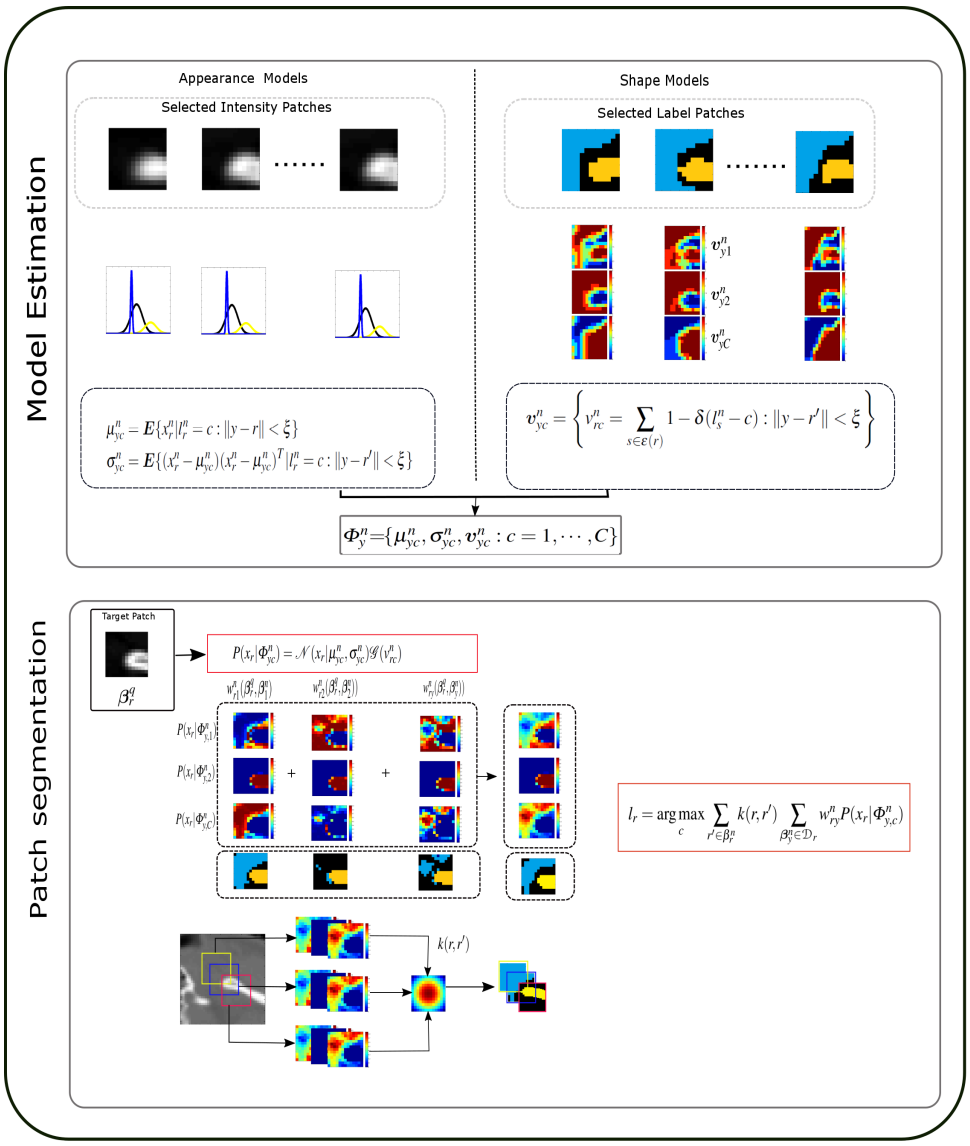


Fig. 2: Proposed segmentation method. For each patch on the dictionary $\mathcal{D}_r$ the appearance and shape model are estimated by a normal and Gibbs distributions respectively. The similarities between the neighboring patches and the target patch are used to measure their contributions. Meanwhile, a gaussian sliding window is used to weights the overlapping neighborhoods estimations based on a the distance to the central voxel.

### 3.2 Image Preprocessing

Each image in the training Dataset, hereafter known as atlas is affine registered to the reference test image using the fiducial registration module of the 3D-Slicer software. The challenge organizers provided the landmarks for the atlases; while for the test ones, they are manually chosen. Finally, all intensity and label atlas images are non-linearly registered to each test image. The registration procedure is performed using the ANTS tool under default parameters: elastic deformation as the mapping function (`Elast`), MI as the similarity metric, and 32-bins histograms for estimating the probability density functions. In order to get a finer alignment, the registration is performed at three sequential resolution levels: *i)* the coarsest alignment with a resolution of $1/8\times$*Original space*, and 100 iterations, *ii)* the middle resolution $1/4\times$*Original space* and 50 iterations, and *iii)* the finest deformation with a resolution of $1/2\times$*Original space* parameter and 25 iterations, the Gaussian regularization method is employed ($\sigma$=3).

### 3.3 Labeling performance

As aforementioned, the testing dataset is composed of two subsets. The off-site images were provided with the training dataset, and the on-site images were provided on the day of the challenge for its segmentation. Only brainstem, parotid glands and mandible were segmented for the challenge. However, the segmentations for all structures were computed and compared with the ground truth labels which were provided by the organizers after the challenge concluding. Segmentation accuracy was measured by Dice similarity coefficient(DSC), mean and standard deviations of dice scores for off-site and on-site subsets are shown in the table Table 1.

| Structures | Off-site | On-site | Average |
|---|---|---|---|
| Brainstem | $0.8626 \pm 0.0406$ | $0.8503 \pm 0.0587$ | $0.8564 \pm 0.0497$ |
| Chiasm | $0.0697 \pm 0.0856$ | $0.0916 \pm 0.0750$ | $0.0806 \pm 0.0803$ |
| Mandible | $0.9386 \pm 0.0264$ | $0.9230 \pm 0.0207$ | $0.9308 \pm 0.0236$ |
| Optic Nerve L | $0.5068 \pm 0.1052$ | $0.4564 \pm 0.1944$ | $0.4816 \pm 0.1498$ |
| Optic Nerve R | $0.5740 \pm 0.0829$ | $0.5382 \pm 0.0292$ | $0.5561 \pm 0.0561$ |
| Parotid L | $0.8123 \pm 0.0425$ | $0.7204 \pm 0.1008$ | $0.7664 \pm 0.0717$ |
| Parotid R | $0.7622 \pm 0.0954$ | $0.7451 \pm 0.1120$ | $0.7536 \pm 0.1037$ |
| Submandibular L | $0.5293 \pm 0.1233$ | $0.5566 \pm 0.1148$ | $0.5430 \pm 0.1191$ |
| Submandibular R | $0.5426 \pm 0.2051$ | $0.3898 \pm 0.1980$ | $0.4662 \pm 0.2016$ |

Table 1: Segmentation results for off-site and on-site datasets

## 4 Discussion and Concluding Remarks

We have presented a novel patch label fusion method based on generative probabilistic approach for weighting the label votes of the neighborhood patches, additionally

a 3D sliding Gaussian window is used to combine the multiples estimations of the overlapping neighborhoods. The proposed method obtained 0.85, 0.93, and 0.76 of average DSC for Brainstem, Mandible, and Parotid glangs respectively. Obtained results are comparable with state of the art methods assessed in [4]. For the other structures the segmentation results were less accurate, this can be due to these exhibit high shape variability and low contrast becoming more complicated its segmentation for patch based approaches.

## References

1. Coupé, P., Manjón, J.V., Fonov, V., Pruessner, J., Robles, M., Collins, D.L.: Patch-based segmentation using expert priors: application to hippocampus and ventricle segmentation. NeuroImage **54**(2) (January 2011) 940–54
2. Zhang, D., Guo, Q., Wu, G., Shen, D.: Sparse Patch-Based Label Fusion for Multi-Atlas Segmentation. MBIA (Mv) (2012) 94–102
3. Wu, G., Wang, Q., Zhang, D., Nie, F., Huang, H., Shen, D.: A generative probability model of joint label fusion for multi-atlas based brain segmentation. Medical image analysis **18**(6) (August 2014) 881–90
4. Sharp, G., Fritscher, K.D., Pekar, V., Peroni, M., Shusharina, N., Veeraraghavan, H., Yang, J.: Vision 20/20: perspectives on automated image segmentation for radiotherapy. Medical physics **41**(5) (2014) 050902