Improving anatomical plausibility in medical image segmentation via hybrid graph neural networks: applications to chest x-ray analysis

Nicolás Gaggion[1,+] Lucas Mansilla[1] Candelaria Mosquera[2] Diego H. Milone[1]
Enzo Ferrante[1]
[1]CONICET / Universidad Nacional del Litoral
[2]Hospital Italiano / Universidad Tecnológica Nacional
[+]ngaggion@sinc.unl.edu.ar

Abstract

Anatomical segmentation is a fundamental task in medical image computing, generally tackled with fully convolutional neural networks which produce dense segmentation masks. These models are often trained with loss functions such as cross-entropy or Dice, which assume pixels to be independent of each other, thus ignoring topological errors and anatomical inconsistencies. We address this limitation by moving from pixel-level to graph representations, which allow to naturally incorporate anatomical constraints by construction. To this end, we introduce HybridGNet, an encoder-decoder neural architecture that leverages standard convolutions for image feature encoding and graph convolutional neural networks (GCNNs) to decode plausible representations of anatomical structures. We also propose a novel image-to-graph skip connection layer which allows localized features to flow from standard convolutional blocks to GCNN blocks, and show that it improves segmentation accuracy.

1. Introduction

Deep convolutional neural networks (CNNs) have achieved outstanding performance in anatomical segmentation of biomedical images. In these images, organs and anatomical structures usually present a characteristic topology that tends to be regular. Since deep segmentation networks are typically trained to minimize pixel-level loss functions, such as cross-entropy or soft Dice [13], their predictions are not guaranteed to incorporate this regularity, due to the inherent lack of sensitivity that these metrics have with respect to global shape and topology [15]. As an alternative, anatomical segmentation can be tackled using other approaches like statistical shape models [9] or graph-based representations [4], which provide a natural way to incorporate topological constraints by construction.

In this work, we explore how landmark-based segmentation can be modeled by combining standard convolutions to encode image features, with generative models based on graph neural networks (GCNNs) to decode anatomically plausible representations of segmented structures.

2. HybridGNet: Image-to-graph extraction via hybrid convolutions

The proposed neural network takes images as input and produces graphs with a fixed number of nodes as output, combining standard convolutions with spectral graph convolutions in a single model that is trained end-to-end. HybridGNet was constructed by combining parts of two independent variational autoencoders (VAE) [11] with the same latent dimension: one to reconstruct images using standard convolutions and another one to reconstruct graphs via spectral convolutions [5, 7]. We decoupled their encoders and decoders, keeping only the image encoder and graph decoder. The HybridGNet was then constructed by connecting these two networks as in Figure 1. This model is trained by minimizing the MSE of the predicted organ contour, under the hypothesis that it directly enforces more anatomically plausible shapes.

Localized image-to-graph skip connections (IGSC) and deep supervision

Under the hypothesis that local image features may help to produce more accurate estimates of landmark positions, we designed a localized Image-to-Graph Skip Connection (IGSC) layer (see Figure 1). IGSC uses the well-known RoIAlign module [8] to sample localized features for each node from a specific encoder level. It receives a tensor of feature maps and a list of node positions which indicate the spatial location from where the feature map will be sampled, and returns the corresponding regions of interest (RoIs) of the given window size centered at the node

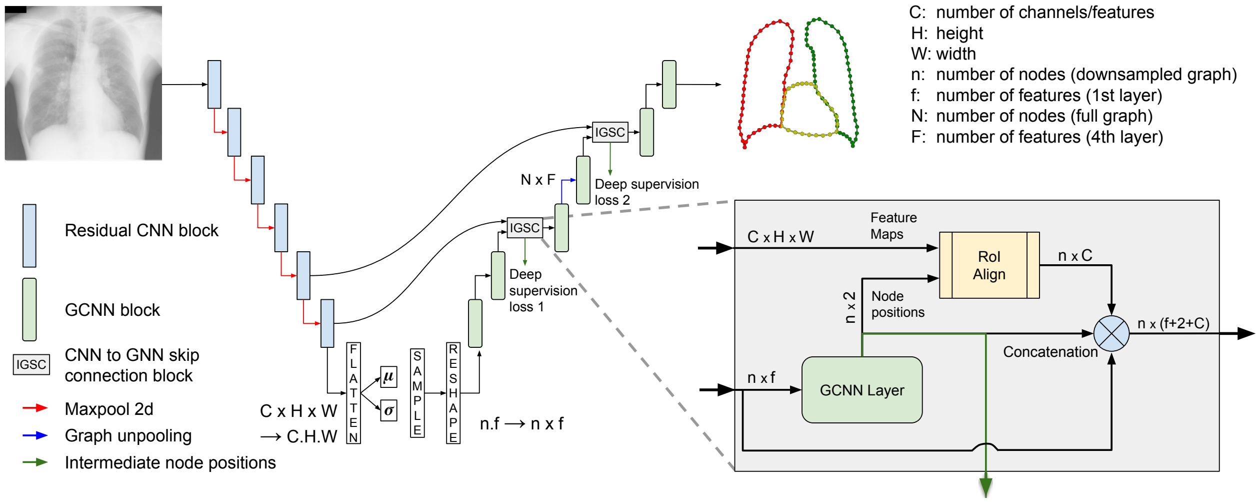

Figure 1. **HybridGNet architecture.** The proposed HybridGNet architecture combines standard convolutions for image feature encoding (blue) with graph spectral convolutions (green) to decode plausible anatomical graph-based representations. The Image-to-Graph skip-connection (IGSC) blocks provide localized features to the intermediate graph representations.

positions. An internal GCNN layer learns the intermediate node positions via deep-supervision [12], resulting in extra loss terms which compute the mean squared error between the ground truth node position and the intermediate predictions. Then, we concatenate the features with the original node features and an augmented graph is obtained. As a hyperparameter, we set the convolutional layer from were we will extract the localized features. We tested our model with 1, 2 or without IGSC.

3. Main experiments

We compared our method with a set of landmark-based segmentation baselines and a pixel-level segmentation model. All except MultiAtlas share the same CNN encoder architecture.

PCA: similar to [3,14], we use PCA to transform the vectorized landmark representation into lower-dimensional embeddings. Then, we optimize the CNN encoder to estimate the PCA coefficients.

FC: We combine the CNN encoder with a fully connected decoder that directly reconstructs the vectorized landmarks.

MultiAtlas: We use a multi-atlas segmentation approach [1, 2], which employs several labeled atlases to delineate the structures of interest.

U-Net [16]: was also included to benchmark our approach against a standard pixel-level segmentation method.

Model comparison

First, we compared HybridGNet with the baselines using the JSRT dataset [17]. We used metrics that can be derived from graph representations, including landmark MSE and Hausdorff distance (HD, in millimeters). To benchmark our methods against the UNet which produces dense segmentation masks, we filled the organ contours to obtain pixel-level masks from graph representations, and computed the Dice coefficient. Table 1 reports metrics on the test set of JSRT dataset (bold numbers indicate significant differences according to Wilcoxon's test). Our HybridGNet model with 2 IGSC outperforms the landmark-based baselines on MSE, Dice, and HD.When compared with the UNet model, HybridGNet surpasses it by a large margin in terms of HD.

Table 1. Landmark-based anatomical segmentation results for JSRT dataset. Mean (std). HD in millimeters.

Model	MSE	Dice Lungs	HD Lungs	Dice Heart	HD Heart
PCA	340.0 (243.6)	0.945 (0.014)	17.44 (9.67)	0.906 (0.037)	14.60 (5.40)
FC	332.2 (242.4)	0.945 (0.017)	17.53 (10.35)	0.910 (0.038)	15.02 (5.78)
MultiAtlas	492.3 (298.1)	0.944 (0.013)	20.32 (9.34)	0.886 (0.056)	16.78 (6.84)
HybridGNet	294.6 (274.5)	0.952 (0.013)	15.64 (10.92)	0.913 (0.038)	13.66 (5.55)
1 IGSC: L6	250.1 (232.0)	0.960 (0.011)	14.38 (9.26)	0.924 (0.030)	**12.34** (4.84)
2 IGSC: L6-5	**200.7** (211.0)	0.974 (0.007)	**12.09** (9.34)	0.933 (0.031)	**11.61** (5.58)
UNet	–	**0.981** (0.008)	21.84 (26.29)	**0.942** (0.030)	25.18 (34.57)

Generating landmark-based representations from dense segmentations

In this work we considered landmark-based segmentations with a fixed number of points, that enable establishing correspondences across images. This is desirable in scenarios like population shape analysis. However, in most segmentation datasets, only pixel-level annotations are available. In these cases, automated estimation of landmarks from dense segmentations can be useful. HybridGNet can be trained to recover landmark-based representations from dense segmentation masks in a natural way. Thus, we trained our best performing models and baselines with dense segmentation masks as input (instead of images), to

Table 2. Results for generating landmark annotations from dense segmentations in the jsrt dataset. Mean (std). HD in millimeters.

Model	MSE	Dice Lungs	HD Lungs	Dice Heart	HD Heart
PCA	77.2 (133.7)	0.978 (0.009)	6.02 (3.46)	0.97 (0.007)	4.37 (1.61)
FC	105.3 (173.2)	0.970 (0.014)	7.82 (3.96)	0.96 (0.014)	5.78 (2.94)
Multi-atlas	236.3 (244.8)	**0.991** (0.004)	10.98 (8.53)	**0.99** (0.006)	4.64 (2.48)
HybridGNet	96.9 (145.0)	0.970 (0.009)	7.65 (3.75)	0.96 (0.013)	6.02 (2.77)
1 IGSC: L6	70.5 (144.9)	0.983 (0.005)	5.54 (5.30)	0.97 (0.011)	4.02 (2.24)
2 IGSC: L6-5	**55.1** (113.4)	**0.991** (0.003)	**3.92** (4.42)	**0.99** (0.005)	**2.58** (1.59)

Table 3. Domain shift results for landmark-based anatomical segmentation from JSRT dataset to Montgomery and Shenzhen. Mean (std). HD in pixels.

	Montgomery		Shenzhen	
Model	Dice Lungs	HD Lungs	Dice Lungs	HD Lungs
PCA	0.906 (0.082)	60.08 (36.89)	0.894 (0.054)	79.12 (47.73)
FC	0.897 (0.087)	60.02 (35.77)	0.895 (0.051)	77.11 (48.15)
Multi-alas	0.909 (0.080)	61.77 (31.62)	0.900 (0.054)	88.13 (48.94)
HybridGNet	0.909 (0.070)	55.97 (35.70)	0.901 (0.047)	72.13 (47.40)
1 IGSC: L6	0.930 (0.062)	48.22 (33.43)	0.914 (0.044)	67.39 (48.53)
2 IGSC: L6-5	**0.954** (0.043)	**45.50** (32.48)	**0.935** (0.038)	**64.46** (51.53)
UNet	**0.944** (0.068)	127.721 (97.76)	0.933 (0.055)	220.89 (102.94)

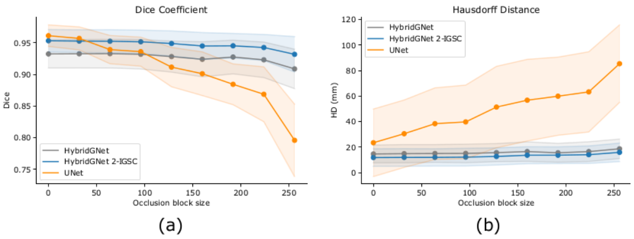

Figure 2. **Artificial occlusions study.** (a) Dice coefficient and (b) HD distance for increasing block size in artificial occlusions.

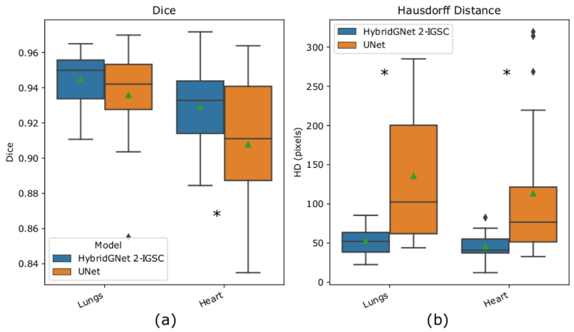

Figure 3. **Real occlusions study.** (a) Dice coefficients and (b) HD distances for the pacemaker Padchest subset.

perform landmark estimation. Table 2 shows the results on the JSRT test set: the proposed HybridGNet with 2 IGSC (Layers 6-5) outperforms the other baselines and architectures. Multi-atlas showed no differences in Dice with respect to our HybridGNet 2 IGSC, but it loses track of the point-to-point correspondences as it is exhibited by the higher MSE error.

Domain shift (DS) evaluation

DS refers to a variation in the target (test) domain concerning the source (training) domain. In most cases, DS drops performance significantly as supervised learning assumes that training samples have the same distribution as the test samples. We compared the effect of DS by measuring segmentation performance on datasets captured at different medical centers, i.e. training in the JSRT dataset and testing with Shenzhen [10] and Montgomery [6] datasets. Table 3 shows how HybridGNet greatly outperform all baselines in terms of HD and Dice coefficient, confirming that the proposed model yields more generalizable predictions across medical centers.

Robustness to image occlusions (IO)

IO are common in chest x-rays, for example due to patient de-identification (i.e., covering protected information with black patches) or external devices such as pacemakers. We designed two experiments to assess the robustness of HybridGNet to artificial and real IO that were not represented in the training set, by comparing it with pixel-level prediction models like UNet.

First, we simulated artificial occlusions by overlapping a random black box on every image. We applied boxes of different sizes over the JSRT test set on random positions. Figure 2 (a) and (b) show Dice and HD distance for lungs and heart segmentation (averaged) as the occlusion block size increases. Although UNet slightly outperforms HybridGNet in Dice for very small boxes, its performance drops with a steeper slope than HybridGNet as we increase the size of the occlusion block.

Robustness to real occlusions produced by external devices was also assessed. To this end, we used 20 segmented images with pacemakers from Padchest as test set. To evaluate solely the occlusion effect on performance and alleviate DS issues, we retrained the models (both HybridGNet and baseline) with an extended training dataset that includes Padchest images (without pacemakers). In Figure 3 we can see how our model outperforms the UNet on both metrics.

4. Conclusions

In this work we present a hybrid network that combines standard CNNs with graph convolutions to decode plausible organ's segmentations. We moved from pixel-level loss functions to a MSE loss function that minimizes the organ's contour distance. We showed that incorporating localized features via image-to-graph skip connections helps to improve the segmentation process greatly. Our HybridGNet produces competitive results on Dice coefficient with respect to models which are trained using soft Dice as their loss function and outperforms them on more topologically aware metrics such as HD.

References

[1] Jennifer Alvén, Fredrik Kahl, Matilda Landgren, Viktor Larsson, and Johannes Ulén. Shape-aware multi-atlas segmentation. In *2016 23rd International Conference on Pattern Recognition (Icpr)*, pages 1101–1106. IEEE, 2016. 2

[2] Jennifer Alvén, Fredrik Kahl, Matilda Landgren, Viktor Larsson, Johannes Ulén, and Olof Enqvist. Shape-aware label fusion for multi-atlas frameworks. *Pattern Recognition Letters*, 124:109–117, 2019. 2

[3] Riddhish Bhalodia, Shireen Y Elhabian, Ladislav Kavan, and Ross T Whitaker. Deepssm: a deep learning framework for statistical shape modeling from raw images. In *International Workshop on Shape in Medical Imaging*, pages 244–257. Springer, 2018. 2

[4] Haithem Boussaid, Iasonas Kokkinos, and Nikos Paragios. Discriminative learning of deformable contour models. In *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pages 624–628. IEEE, 2014. 1

[5] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and locally connected networks on graphs. *arXiv*, 2013. 1

[6] Sema Candemir, Stefan Jaeger, Kannappan Palaniappan, Jonathan P. Musco, Rahul K. Singh, Zhiyun Xue, Alexandros Karargyris, Sameer Antani, George Thoma, and Clement J. McDonald. Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration. *IEEE Transactions on Medical Imaging*, 33(2):577–590, 2014. 3

[7] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *arXiv preprint arXiv:1606.09375*, 2016. 1

[8] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. 1

[9] Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3d medical image segmentation: a review. *Medical image analysis*, 13(4):543–563, 2009. 1

[10] Stefan Jaeger, Alexandros Karargyris, Sema Candemir, Les Folio, Jenifer Siegelman, Fiona Callaghan, Zhiyun Xue, Kannappan Palaniappan, Rahul K. Singh, Sameer Antani, George Thoma, Yi-Xiang Wang, Pu-Xuan Lu, and Clement J. McDonald. Automatic tuberculosis screening using chest radiographs. *IEEE Transactions on Medical Imaging*, 33(2):233–245, 2014. 3

[11] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013. 1

[12] Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, and Zhuowen Tu. Deeply-supervised nets. In *Artificial intelligence and statistics*, pages 562–570, 2015. 2

[13] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. IEEE, 2016. 1

[14] Fausto Milletari, Alex Rothberg, Jimmy Jia, and Michal Sofka. Integrating statistical prior knowledge into convolutional neural networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 161–168. Springer, 2017. 2

[15] Annika Reinke, Matthias Eisenmann, Minu D. Tizabi, Carole H. Sudre, Tim Rädsch, Michela Antonelli, Tal Arbel, Spyridon Bakas, M. Jorge Cardoso, Veronika Cheplygina, Keyvan Farahani, Ben Glocker, Doreen Heckmann-Nötzel, Fabian Isensee, Pierre Jannin, Charles E. Kahn, Jens Kleesiek, Tahsin Kurc, Michal Kozubek, Bennett A. Landman, Geert Litjens, Klaus Maier-Hein, Bjoern Menze, Henning Müller, Jens Petersen, Mauricio Reyes, Nicola Rieke, Bram Stieltjes, Ronald M. Summers, Sotirios A. Tsaftaris, Bram van Ginneken, Annette Kopp-Schneider, Paul Jäger, and Lena Maier-Hein. Common limitations of image processing metrics: A picture story, 2021. 1

[16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 2

[17] Junji Shiraishi, Shigehiko Katsuragawa, Junpei Ikezoe, Tsuneo Matsumoto, Takeshi Kobayashi, Ken-ichi Komatsu, Mitate Matsui, Hiroshi Fujita, Yoshie Kodera, and Kunio Doi. Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists' detection of pulmonary nodules. *American Journal of Roentgenology*, 174(1):71–74, 2000. 2