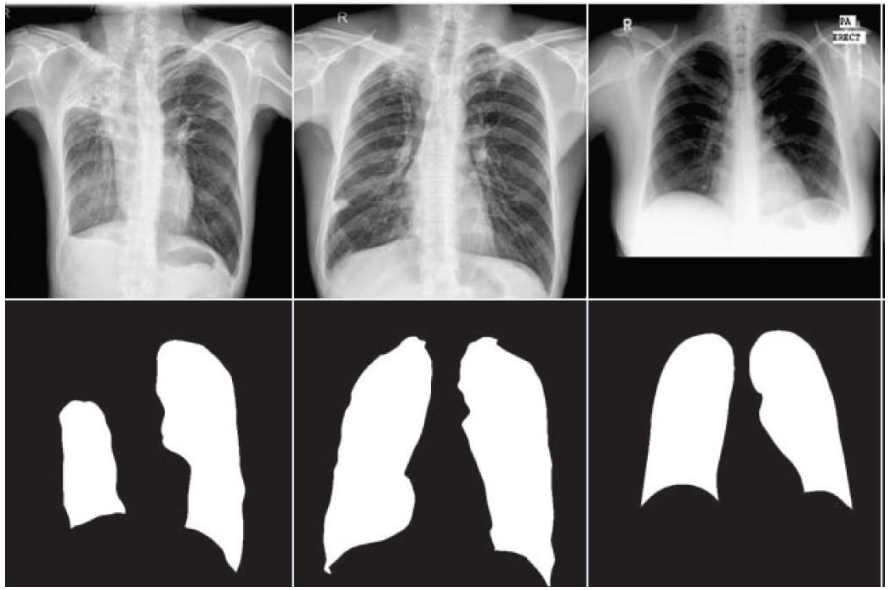
Robust Boundary Segmentation in Medical Images Using a Consecutive Deep Encoder-Decoder Network

Ngoc-Quang Nguyen; Sang-Woong Lee



Supervisor: Sang-Woong Lee

First-author: Ngoc-Quang Nguyen

PRML

CONTENTS

- Contributions
- Methods
  - Boundary-Emphasization Data Augmentation
  - Consecutive deep encoder-decoder network
- Datasets
- Experimental results
- Future works
- Q&A

CONTRIBUTIONS

- We propose a new continuous multiple deep encoder-decoder network, CDED-Net, to extract the most useful features from images and learn completely from multiscale image inputs.

- We introduce a boundary-emphasization augmentation method for making a high number of object boundary patterns from each image in a training set. The novel augmentation method enhances and boosts the segmentation performance of CDED-net.

- In our CDED-net, instated of using constraint dilated convolution, we use different both of strides and rates for each component network to capture contextual information at multiple scales input.

- We present a new Dicoss-loss function, which is a measure of overlap widely used to assess segmentation performance of a network. The combination of the loss function and our CDED-net results in a better performance.

METHODS

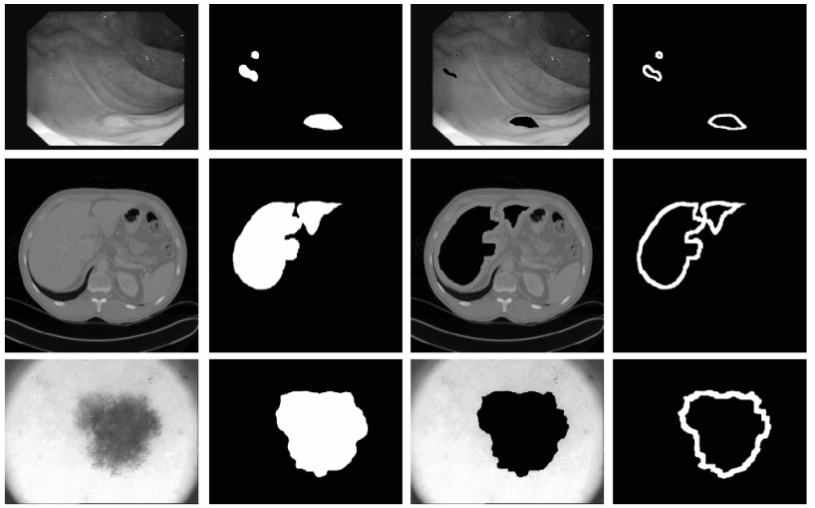
Boundary-Emphasization Data Augmentation

Data augmentation is the creation of altered copies of each instance within a training dataset to increase the number of images.

Owing to its medical characteristics, the non-specular transition zone between the medical object and its surrounding area is not easy to discriminate. This area does not differ dramatically from other areas.

Algorithm 1 Boundary-Emphasization Augmentation Algorithm

Input: Input image I, corresponding label I_L;
Object region S;
Structuring element (erosion kernel size) C;
Erosion represented by $\ominus$;
Euclidean space E;
Erosion label I_L^E;
C_z is the translation of C by the vector z,
$\forall\, z \in E$;
Processed label I_L^*;
Output: Processed image I^*, corresponding processed label I_L^*
while *True* **do**
$I_L^E = I_L \ominus C = \{z \in E \mid C_z \subseteq I_L\}$;
$I_L^* = I_L$ - I_L^E
$i(x_I, y_I) \subset (I)$;
if $i^*(x_I, y_I) \subset I_L$ **and** $i^*(x_I, y_I) \subset S_{I_L^E}$ **then**:
$i = 0$;
else:
pass;
result $I^* \leftarrow I, I_L^* \leftarrow I_L$;



a) b) c) d)

METHODS

CONSECUTIVE DEEP ENCODER-DECODER NETWORK

- The advantages and disadvantage of DeepLabV3+ inspired us to develop the proposed network.
- The objective of this study is to build an ensemble of deep encoder-decoder networks to train and obtain rich contextual information for the medical object segmentation task.
- To capture contextual information at multiple scales, we used deep encoder-decoder networks, namely DeepLab V3X

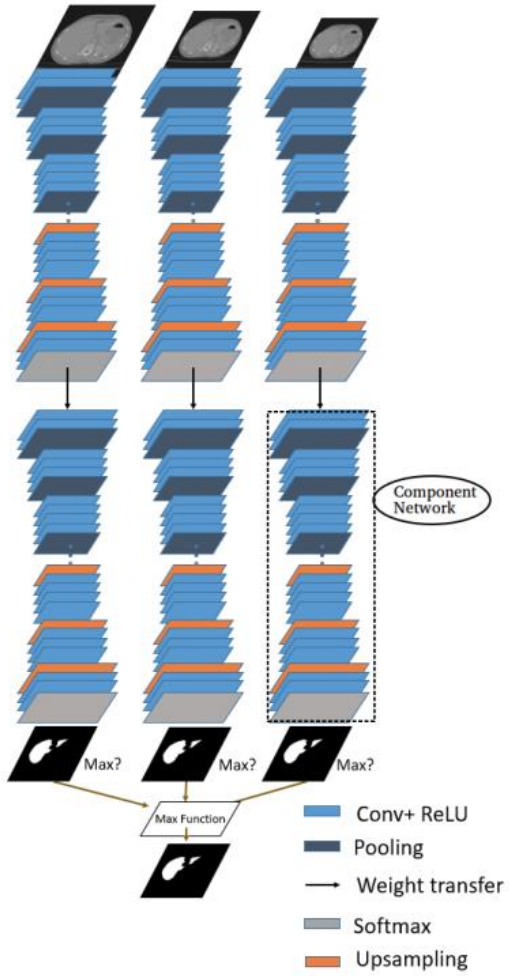


FIGURE 3. The entire architecture of proposed CDED-net.

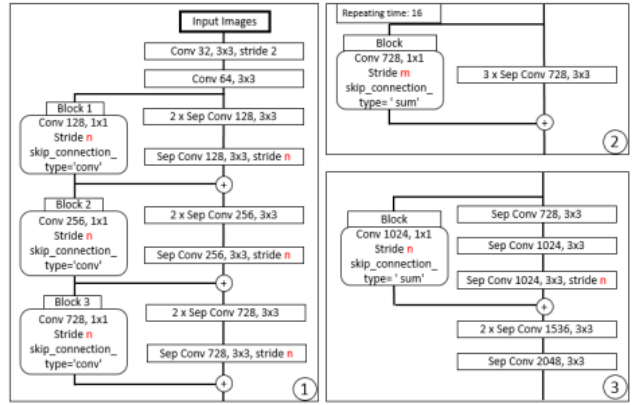


FIGURE 4. The detail of the component proposed CDED-net with Deeplab V3+ [11] as the backbone. 1) Entry flow. 2) Middle flow. 3) Exit flow. We modified the convolution stride to adapt with the resolution of dataset for extracting rich information features.

METHODS

DICOSS LOSS FUNCTION

- The loss simply verified each pixel individually, comparing the class predictions that are defined as depth-wise pixel vector to the target vector.
- The loss function's pixel-level assessment can be problematic with multiple classes in an image, especially in medical images with limited surface area.
- This biases the segmentation network towards the background over the object. Combining it with a dice loss function mitigates these issues.

$$L_c = -(1-\gamma) * \sum_{i,j} \hat{y}_{i,j} * log(y_{i,j}) + \gamma \frac{2\sum_{i,j} \hat{y}_{i,j} * y_{i,j}}{\sum_{i,j} \hat{y}_{i,j}^2 + \sum_{i,j}^{N} y_{i,j}^2} \quad (3)$$

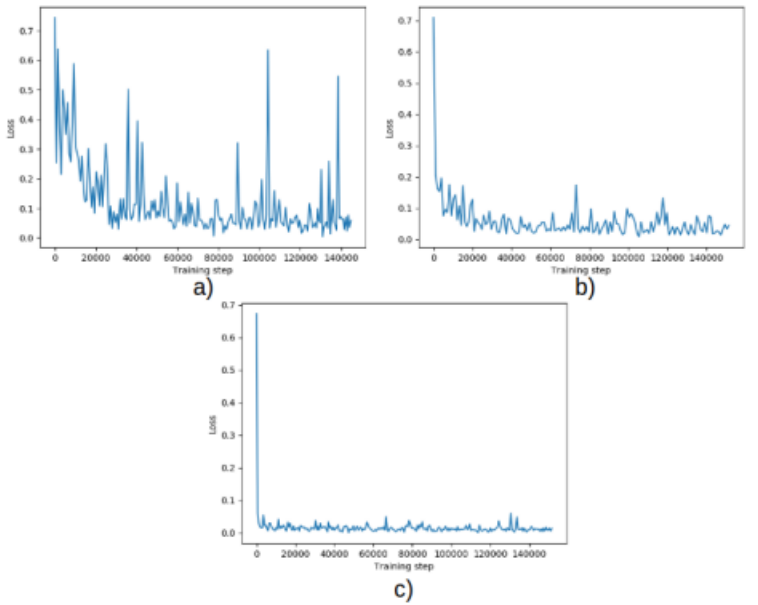


FIGURE 7. The effect of proposed loss function to network learning progress on the same dataset by comparing to two fundamental loss functions. a) Basic cross entropy loss function. b) Basic Dice loss function. c) Proposed loss function.

DATASETS

• CVC-ClinicDB [37] contains 612 images, where all images show at least one polyp. The segmentation labels obtained from 31 colorectal video sequences were acquired from 23 patients.

• CVC-ColonDB [46] ontains 379 frames from 15 different colonoscopy sequences, where each sequence shows at least one polyp each.

• ETIS-LaribPolypDB [45] contains 196 images, where all images show at least one polyp.

• PH2 [38] contains 200 dermoscopic images with a resolution of 768 × 560 pixels that were acquired at Dermatology Service of Hospital Pedro Hispano, Matosinhos, Portugal Mendonça with Tuebinger Mole Analyzer system, this dataset includes 80 common nevus images, 80 atyp-ical nevus images and 40 melanoma image

• ISBI 2016 [56] contains 900 training images with the ground truth provided by experts. The image sizes vary from 1022 × 767 to 4288 × 2848 pixel. This dataset was provided at the 2016 International Symposium on Biomedical Imaging (ISBI 2016).

• CHAOS 2019 [57] contains 980 liver CT images with re resolution is 512 × 512 pixel in DICOM format. This dataset was provided at the IEEE International Symposium on Biomedical Imaging (ISBI) on April 8-11, 2019.

EXPERIMENTAL SETTINGS

- Novel pair (Davis): There were no overlaps between the training and test datasets. Neither the training compound nor the training protein appeared in the test set.
- Novel compound (Davis): There were no intersections of compounds in the training set and compounds in the test set.
- Novel protein (Davis): There were no intersections of proteins in the training set and proteins in the test set.
- Novel hard pair (Metz, KIBA): We removed interactions from the training dataset if either the protein sequence or the compound had a similarity score exceeding the threshold 0.3
- Cross-domain (Metz, CASF-2016): We removed interactions involving 56 proteins and 105 compounds with similarities higher than 0.3 from the Metz dataset.
- Enrichment factor analysis (BingdingDB, DUD_E diverse): we removed interactions for two proteins and compounds that appeared in both datasets (GCR HUMAN (P04150) and AKT1 HUMAN (P31749) and 102 compounds) from training set.

EXPERIMENTAL RESULTS

TABLE 1. Comparison of proposed method and three fully convolutional neural networks in terms of mean pixel precision and recall for the ETIS-Larib dataset [45].

Networks	Mean pixel precision	Mean pixel recall
FCN-AlexNet [20]	0.2789	0.3554
FCN-GoogLeNet [21]	0.2583	0.3782
FCN-VGG [22]	0.7023	0.5420
Proposed	**0.9293**	**0.9087**

TABLE 2. Comparison of proposed method with FCN-8S combined with post-processing and a combination of fully convolutional neural network and textons on CVC-ColonDB dataset [46].

Networks	Accuracy	Specificity	Dice	Sensitivity
Zhang et al. [19]	0.975	0.988	0.701	0.757
Akbari et al. [6]	0.977	**0.993**	0.810	0.748
Proposed	**0.980**	0.991	**0.896**	**0.792**

TABLE 3. Comparison of proposed method on CVC-ClinicDB dataset [37].

Networks	Accuracy	Specificity	Sensitivity	Dice	Precision
Li et al [24]	0.97	0.77	**0.99**	0.83	0.90
Proposed	**0.987**	**0.942**	0.962	**0.891**	**0.950**

TABLE 4. Comparison of proposed method on PH^2 dataset [38].

Networks	IoU	Dice	Sensitivity	Specificity	Accuracy
MGAC [25]	87.03	92.79	93.59	97.81	**96.18**
FCN [5]	82.59	90.46	95.35	94.09	94.44
U-Net [8]	81.63	89.88	86.68	97.63	86.68
SegNet [16]	84.03	91.32	91.57	96.57	95.19
FrCN [49]	84.13	91.38	94.48	95.46	95.20
MSCA [52]	76.88	85.52	85.78	96.33	93.86
MFCN [53]	84.15	90.77	95.64	95.12	95.61
DCL-PSI [51]	86.05	92.26	**97.11**	95.85	96.61
Tong et al. [48]	60.0	75.0	-	-	-
DermoNet [26]	85.3	91.5	-	-	-
Proposed	**88.78**	**94.10**	96.23	**97.84**	95.40

EXPERIMENTAL RESULTS

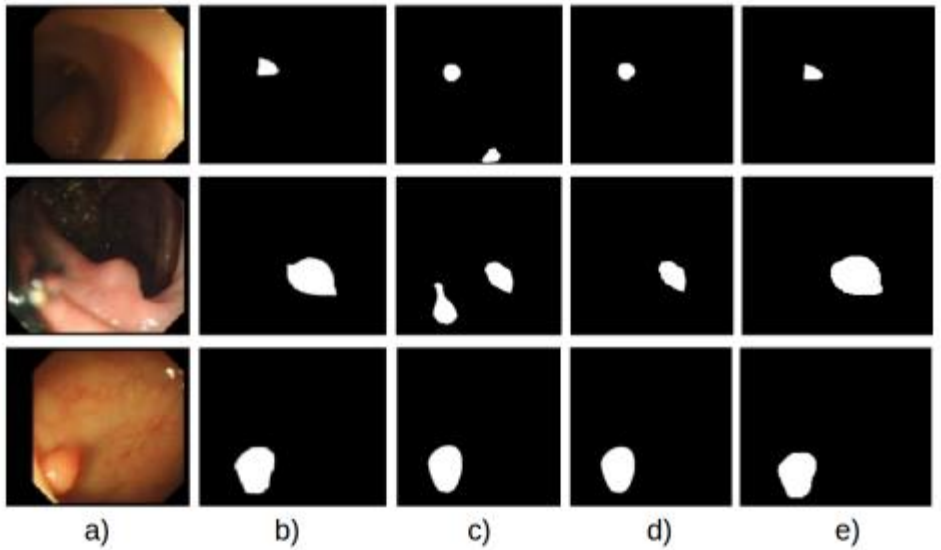


FIGURE 8. Comparison of proposed method with [6]. a) Input images. b) Ground truth. c) FCN-8S with Otsu threshold. d) FCN-8S final result. e) Proposed method.

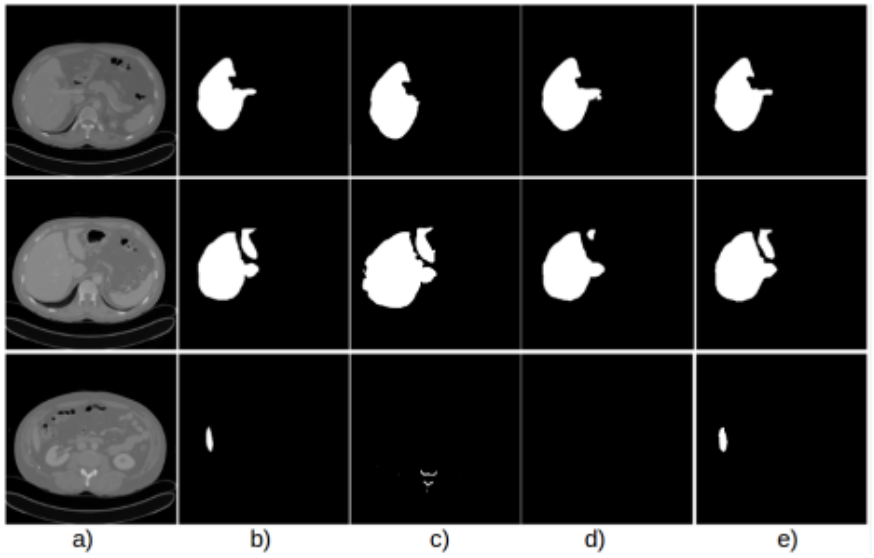


FIGURE 9. Comparison of proposed method on CHAOS dataset [57]. a) Input images. b) Ground truth. c) U-net [8]. d) Deeplab V3+ [11]. e) Proposed method.

FUTURE WORKS

- incremental boosting convolution networks by adopting other novel effective methods such as using the advantages of a neural architecture search (NAS) algorithm that can support the network:
  - allowing it to focus on searching the repeatable cell structure, while hand designing the outer network structure that controls the spatial resolution changes.

THANK YOU FOR YOUR ATTENTION

Q&A