# How loss function affects multiclass Deeplab segmentation of abdominal organs MRI

Pedro, N., Furtado*
Universidade de Coimbra, CISUC, DEI

## ABSTRACT

Loss function is the fundamental driver of backpropagation learning in deep convolutional neural networks (DCNN). There exist alternative formulations, such as cross entropy, jaccard and dice. But does the choice of loss influence quality decisively? Another relevant question is whether the quality of the best approaches nears perfection, as some seem to suggest. In this paper we investigate how variations of loss function affect the outcome of segmentation of abdominal organs by Magnetic resonance imaging (MRI). Variations include changing the loss function (cross entropy, dice, IoU), but also penalizing differently false positives and false negatives. We conclude that, for this problem of segmentation of abdominal organs, we were able to improve the quality of segmentation by 6% based on tuning the loss function on the best performing segmentation network (deeplabV3). The conclusions are important for anyone trying to segment these or other organs and structures. Future work is required to generalize the conclusions on other datasets as well, and to conclude in what concerns factors that determine the best choice.

## CCS CONCEPTS

• **Applied computing**; • **Life and medical sciences**; • **Bioinformatics**;

## KEYWORDS

Computers in Medicine, Segmentation, Machine Learning, Deep Learning, MRI

**ACM Reference Format:**
Pedro, N., Furtado. 2021. How loss function affects multiclass Deeplab segmentation of abdominal organs MRI. In *2021 13th International Conference on Bioinformatics and Biomedical Technology (ICBBT '21), May 21–23, 2021, Xi'an, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3473258.3473260

## 1 INTRODUCTION

Magnetic resonance imaging (MRI) is an imaging technology based on applying strong magnetic fields to affect proton spin movements, then releasing the magnetic field and observing specific relaxation times (the return of protons to the original resting phase). An MRI scan produces a sequence of two-dimensional slices, allowing a full study and 3D reconstruction of the body structures captured in those slices. Segmentation is an algorithmic procedure which further individualizes specific organs and structures. While manually tuned segmentation algorithms, based on features extraction and atlas information were the norm some years ago, deep learning revolutionized the procedure by introducing automatic end-to-end learning from training images and groundtruths. In those systems a large number of iterations improves the quality of the output sequentially.

In general, deep learning has since been used to both classify and segment all kinds of medical images. Examples of segmentation applied to Magnetic resonance images (MRI) include acute ischemic lesions [1], brain tumors [2], the striatum [3], organs-at-risks in head and neck [4], polycystic kidneys [5], prostate [6] and spine [7], applications also being reviewed in [8] and [9].

The principal objective of the segmentation algorithm in this context is to achieve as good as possible delineation of the organs and structures that are targeted by the approach. Given a set of training images and groundtruth segments, the deep learning segmentation system trains based on a set of iterations. On each iteration it takes each image sequentially through a set of convolution and deconvolution stages, which operate mathematical operations on the inputs based on coefficients, and outputs a segmentation mask. That segmentation mask is compared with the groundtruth masks (the correct segmentation) using a function that is called the loss function, since it computes the loss or error. The loss is then backpropagated to iteratively modify the coefficients slightly along the whole network in each iteration, using a gradient descent algorithm or similar. The learning process is therefore a large number of iterations of this backpropagation learning, until at the end one expects the loss to be minimized and the segmentation to be as good as possible. It is however difficult for the loss function to be perfect. In the next example of Figure 1 the final loss of the training procedure was less than 2%, yet we can see an example of incorrect segmentation.

Given the importance of the loss function in the training of the networks, perhaps its choice has definite influence in the quality of the results? Another important question is how perfect is segmentation using as good as possible loss function? For these reasons, it is a relevant topic to experiment with different formulations. However, it is also important to understand why there are limitations with the loss function, which leads us to study metrics. Loss is a metric, exactly as the metrics used to evaluate quality of segmentation. We need to understand the limitations of metrics themselves, and then the limitations of loss as a metric, in order to understand why a solution with very small loss might in fact be visually imperfect.

In this paper we first study the metrics and loss alternatives, highlighting the problems exhibited by each. Afterwards, we resort to an experimental approach to study the effect of different loss functions and variations on the quality of segmentation of MRI

*Place the footnote text for the author (if applicable) here.



*ICBBT '21, May 21–23, 2021, Xi'an, China*

ACM ISBN 978-1-4503-8965-5/21/05.
https://doi.org/10.1145/3473258.3473260

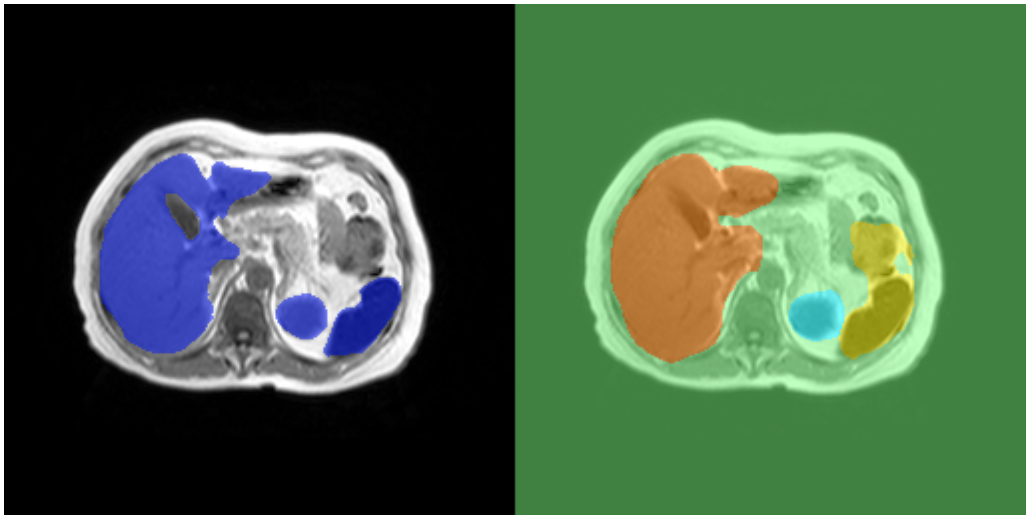

**Figure 1: MRI segmentation of independent test images using DeepLabv3. The left shows the groundtruth, the right shows the segmentation. It is visible that the segmentation is still not perfect.**

sequences of abdominal organs. This allows us to better understand the effect of loss in that context. We achieve a 6% improvement over the default cross entropy loss function. We also define as future challenge experimenting with other datasets and arriving at a set of rules regarding the choice of best loss function depending on the data. These results extend prior work in the detail of evaluating how effective is a choice of loss function and some variations that can be introduced. It also focuses on quantifying the quality achievable by a choice of network and loss function, in the context of MRI of abdominal organs.

## 2 RELATED WORK

The segmentation network is a modified DCNN architecture that classifies each pixel (with a segment label) instead of the image. To achieve this, the fully connected layers are replaced by a decoder that successively de-convolves until the full image size is restored. The Fully Convolutional Network (FCN) [10] was one of the first well-structured segmentation network architectures. It uses a DCNN as encoder (e.g. VGG) and replaces the final fully-connected layers by up-sampling interpolation layers. U-Net introduced further innovations [13], with de-convolution stages symmetric to the convolution stages (forming a U-shape) instead of interpolation. De-convolutions combine feature and spatial information through a sequence of up-convolutions and concatenations with high-resolution features from the contracting path. DeepLab [12] is another highly accurate segmentation network that introduces important innovations. One such innovation is Atrous Spatial Pyramid Pooling (ASPP), which improves segmentation at multiple scales. Another innovation is the use of Conditional Random Fields (CRF) that apply probabilistic graphical models for improved determination of objects boundaries.

Recent works on segmentation of MRI and CT (computer tomography) images include Zhou [13], which used a fully convolutional networks (FCN) applying a majority voting scheme on the output of segmentation of 2D slices taken from different image orientations of CT. Then [14] applied a similar approach to segmentation of the abdomen from MRI sequences, Larsson [15] proposed SeepSeg and [16] proposed multi-slice 2D neural network designed in a way that considers information of subsequent slices, plus augmented data and multiview training. Groza [17] presents an ensemble of DL networks with voting, and [18] tests different architectures (U-Net, deeper U-Net with VGG-19, a cascade of two networks). In [19] the authors propose a new MRI segmentation method using a CNN-based correction network for MRI-guided adaptive radiotherapy, having achieved high segmentation scores. In [20], Chlebus et al. studied reducing inter-observer variability and interaction time of MR liver volumetry by combining automatic CNN-based liver segmentation and manual corrections, achieving jaccard segmentation score of 0.9 for the liver. In [21] the authors proposed automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets. [22] proposes an organ-attention networks and statistical fusion, and [23] proposes a multi-scale pyramid of 3D fully convolutional networks for abdominal multi-organ segmentation.

Compared to all those works we pose a different question, namely how useful are loss function variations and how perfect is segmentation with the best performing one. Although loss has not been the centerpiece in approaches to improve segmentation quality, some authors pay some attention to it. In particular, [24] proposed improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and "direct" loss function. They propose a Jaccard Loss. [18] also replaced cross-entropy but by the dice function to better deal with class imbalance. The loss function is based on a metric, and the problem of metrics in general is mentioned in [25]: "many scores are artificially high simply because the background is huge and hence the term TN (true negatives) is also huge". These works have shown the importance of loss and the limitations of metrics in general. In this paper we take these

lessons to propose and experiment with different loss alternatives in the context of MRI of abdominal organs.

# 3 MATERIALS AND METHODS

## 3.1 Influence of metrics and loss

In most bibliography, metrics are defined considering a binary classification problem that classifies into two classes: positive (P), with the meaning "is", and negative (N), with the meaning "is not". The quantities TP, TN, FP and FN correspond to the number of pixels that are true positives, true negatives, false positives and false negatives, respectively. Given those quantities, some of the most frequently used metrics are:

$$\text{Accuracy (ac)} = (\text{TP} + \text{TN}) \,//\, \text{TP} + \text{TN} + \text{FP} + \text{FN}); \quad (1)$$

$$\text{Sensitivity (se)} = \text{recall} = \text{TruePositiveRate (TPR)} = \text{TP}/\,(\text{TP} + \text{FN}) \quad (2)$$

$$\text{Specificity (sp)} = \text{TN}/\,(\text{TN} + \text{FP}) \quad (3)$$

$$\text{Precision (p)} = \text{TP}/\,(\text{TP} + \text{FP}) \quad (4)$$

$$\text{False Positive Rate (fpr)} = \text{FP}/\,(\text{FP} + \text{TN}) \quad (5)$$

$$\text{ROC, a plot of TPR vs FPR, and AUC, the area under the curve of ROC} \quad (6)$$

$$\text{IoU} = \text{JI} = \text{TP}/\,(\text{TP} + \text{FN} + \text{FP}) \quad (7)$$

$$\text{Dice (dice)} = \text{DSC} = 2\text{TP}/\,(2\text{TP} + \text{FP} + \text{FN}) = 2\text{JI}/\,(\text{JI} + 1) \text{, which is highly correlated with JI} \quad (8)$$

In multiclass problems we can apply the same formulas, but considering the following quantities instead: a TP pixel is a pixel that belongs to one class c different from background in groundtruth and also in the segmentation; a TN pixel is a pixel that belongs to background in both groundtruth and segmentation; an FP pixel is a pixel that belongs to background in groundtruth but is classified as some other class c in segmentation; an FN pixel is a pixel that belongs to some class c different from background in groundtruth but is then classified as background;

The following three observations are important reasons why the metrics defined in equations (1) to (8) can fail to evaluate segmentation correctly in many medical imaging contexts:

a) the number TP is always huge in all metrics, because TP of background pixels is huge. As a consequence, all metrics (1) to (8) report high scores regardless of the actual quality of segmentation of individual organs if evaluated over all pixels;
b) TN is also huge because it includes a huge number of background pixels that are well classified. It means that specificity (SP), FPR, ROC and AUC do not evaluate the quality of segmentation of organs well;
c) Sensitivity (a.k.a recall or TPR), although useful because it quantifies the fraction of organ pixels classified correctly as such, fails to capture very important possible deficiencies, because it does not include FP (background classified as organ) in the formula, a frequent occurrence.

Given the previous observations we conclude that metrics need to be evaluated and reported separately for each class (TP, TN, FP and FN replaced by TPc, TNc, FPc and FNc, where c is a class).

But the loss function, which is also a metric, needs to output a single value to be used as delta in backpropagation learning, therefore it must be averaged over the loss of each class, in spite of the fact that there is a huge imbalance with class background. That is the reason why the training and validation losses can be so low in a quite imperfect segmentation.

Based on the previous discussion we define a set of loss functions besides cross entropy, and a set of variations:

**Cross entropy (crossE, the default to compare with):** cross entropy measures dissimilarity between pixel classification and actual pixel class. If ti and si are the groundtruth and the CNN score of each pixel for each class i respectively,

$$crossE = -\sum_{i}^{C} t_i \log s_i \quad (9)$$

**Intersect over the Union (IoU or JI):**

$$IoU\,(loss) = 1 - IoU = 1 - \frac{TP}{TP + FP + FN} \quad (10)$$

IoU over classes is ,

$$IoU\,(loss) = 1 - \frac{\sum_{I=1}^{C} IoU_i}{C}, IoU_i = 1 - \frac{TP_i}{TP_i + FP_i + FN_i} \quad (11)$$

Dice (dice): The dice (or DSC) is highly correlated to IoU,

$$dice\,(loss) = 1 - DSC = 1 - \frac{2TP}{2TP + FP + FN} \quad (12)$$

$$dice\,(loss) = 1 - \frac{\sum_{I=1}^{C} dice_i}{C}, dice_i = 1 - \frac{2TP_i}{2TP_i + FP_i + FN_i} \quad (13)$$

**Intersect over the Union with penalties (IoUxy): IoUxy is similar to IoU but penalizes differently FP and FN in the denominator of the formula:**

$$IoU_{xy}\,(loss) = 1 - \frac{\sum_{I=1}^{C} IoU_{xyi}}{C}, IoU_{xyi} = 1 - \frac{TP_i}{TP_i + \alpha FP_i + \beta FN_i} \quad (14)$$

In these formulas $\alpha$ and $\beta$ are such that $\alpha$+$\beta$ =2, $\alpha, \beta$>=0.

## 3.2 dataset and experimental setup

The Magnetic Resonance Imaging data used in our experiments is a set of scans available in [26]. It consists of (MRI) acquisitions of 120 MRI sequences from 120 healthy patients (routine scans, no tumors, lesions or any other diseases), capturing abdominal organs (liver, kidneys and spleen), obtained using T1-DUAL fat suppression protocol. The sequences were acquired by a 1.5T Philips MRI, which produces 12-bit DICOM images with a resolution of 256 x 256. The ISDs varies between 5.5-9 mm (average 7.84 mm), x-y spacing is between 1.36 - 1.89 mm (average 1.61 mm) and the number of slices per scan is between 26 and 50 (average 36). Train, test and validation data independent form each other were always obtained by dividing the patients into those subsets. The proportion was 80%/20% train/test. Data augmentation was defined based on random translations of up to 10 pixels, random rotations up to 10 degrees, shearing up to 10 pixels and scaling up to 10%.

**Table 1: IoU of segmentation networks with base crossE loss.**

| class | DeepLabV3 | FCN | UNET |
|---|---|---|---|
| **Background** | 99% | 99% | 98% |
| **Liver** | 86% | 86% | 74% |
| **Spleen** | 82% | 74% | 73% |
| **rKidney** | 77% | 78% | 75% |
| **lKidney** | 81% | 77% | 78% |
| **Avg IoU** | **85%** | **83%** | **80%** |

We experimented with U-NET [11], FCN [10] and DeepLabV3 [12]. The experiments reported in this work were preceded by a set of iterations tuning configurations to the best possible results. The final network training parameters after tuning, to be used in our experimental work, were: learning function Stochastic gradient descent with momentum (SGDM), with an initial learning rate=0.005, piecewise learning rate with drop period of 20 and learn rate drop factor of 0.9 (i.e. the learn rate would decrease to 90% every 20 epochs). training iterations were 500 epochs; minibatch sz=32; momentum= 0.9. The factor that most improved performance in our initial tuning prior to experiments was data augmentation, which we described before. A machine with a GPU NVIDEA G Force GTX1070 was used for the experiments.

# 4 RESULTS

## 4.1 Choose best-performing network

Table 1 shows IoU (JI) of UNet, FCN and DeeplabV3 with crossE loss. The best-performing network was DeepLabV3 (85% vag IoU versus 80 to 83%). Our next experiments use that network.

## 4.2 Comparison of loss function variations

Table 2 shows the global scores of the different loss functions, and table 3 details the results further by displaying loss scores for each organ measured as IoU.

Given that IoU with different weights to false positives and false negatives was superior to other choices, we also experimented with varying the alpha coefficient in the denominator of IoU loss function ($\alpha FP_i + \beta FN_i$).

# 5 DISCUSSION

In table 2 accuracy and weighted IoU always scored very high, while the remaining metrics better reveal deficiencies and allow us to compare the approaches better. IoU in table 2 shows that the best performing loss function was IoU0515, scoring 0.9 and improving 6 percentage points (pp) when compared with the default cross entropy (crossE) loss function (and 6 pp on sensitivity). Most alternatives (i.e. crossE, IoU11, dice, dice no BK) had similar average scores (IoU and dice are highly correlated metrics). Table 3 shows the detail for each organ. IoU0515 achieved scores between 0.85 and 1 for the different classes. We conclude that IoU loss with modified weights on FN and FP can improve segmentation quality (IoU0515 assigns a weight of 25% (0.5/2) to false positives (FP) and 75% weight (1.5/2) to false negatives (FN)). Dice was the second best alternative. The final conclusion from these observations is that IoU or dice should be used to improve scores, but also that it is worth experimenting with different combinations of IoU weights.

**Table 3: IoU of segmentation network DeepLabV3 with diff. loss functions.**

| IoU | crossE | IoU | IoU | Iou | dice |
|---|---|---|---|---|---|
| $\alpha$ | - | 1 | 1.5 | 0.5 | - |
| $\beta$ | | 1 | 0.5 | 1.5 | |
| **BackGround** | 0.99 | 0.99 | 0.99 | **1.00** | 0.99 |
| **liver** | 0.86 | 0.84 | 0.69 | **0.88** | 0.87 |
| **spleen** | 0.82 | 0.84 | 0.80 | **0.87** | 0.80 |
| **rkidney** | 0.77 | 0.82 | 0.77 | **0.88** | 0.81 |
| **lkidney** | 0.81 | 0.74 | 0.73 | **0.85** | 0.76 |
| **avg** | **0.84** | **0.86** | **0.82** | **0.90** | **0.85** |
| **rank** | 3 | 2 | 9 | 1 | 3 |

## 5.1 Results by other works

Although not focusing on loss function, some other authors also ran segmentation on the same dataset that we used. Table 4 shows scores of a few other approaches running on the same dataset as ours (therefore directly comparable). Our best performing alternative was superior to those compared in the table.

Finally, Figure 2 shows visualizations of segmentation results with our DeepLabV3 network, where we can see that the segmentations succeed at finding the organs areas, although with some imperfections.

# 6 CONCLUSIONS

In this work we have investigated loss function variations for the problem of magnetic Resonance imaging (MRI) of abdominal organs. We have argued that the loss function is fundamental due

**Table 2: Global metrics for segmentation network DeepLabV3 with base crossE loss.**

| | Accuracy | Mean Sensitivity | Mean IoU | Weighted IoU | Mean BFScore |
|---|---|---|---|---|---|
| crossE | 0.99 | 0.88 | 0.84 | 0.99 | 0.90 |
| iou11 | 0.99 | 0.88 | 0.86 | 0.99 | 0.90 |
| iou1505 | 0.99 | 0.83 | 0.82 | 0.98 | 0.85 |
| **iou0515** | **1.00** | **0.94** | **0.90** | **0.99** | **0.92** |
| dice | 0.99 | 0.87 | 0.85 | 0.99 | 0.91 |
| **dice noBK** | 0.99 | 0.89 | 0.85 | 0.99 | 0.90 |

**Table 4: Comparing to IoU of related approaches (CHAOS dataset).**

| MRI JI=IoU | Liver | spleen | R Kidney | L kidney |
|---|---|---|---|---|
| [18] teamPK | | | | |
| U-Net | 0.73 | 0.76 | 0.79 | 0.83 |
| V19UNet | 0.76 | 0.79 | 0.84 | 0.85 |
| V19pUNet | 0.85 | 0.83 | 0.85 | 0.86 |
| V19pUnet1-1 | 0.86 | 0.83 | 0.86 | **0.87** |
| deeplabV3 iou 0.5/1.5 | **0.88** | **0.87** | **0.88** | 0.85 |

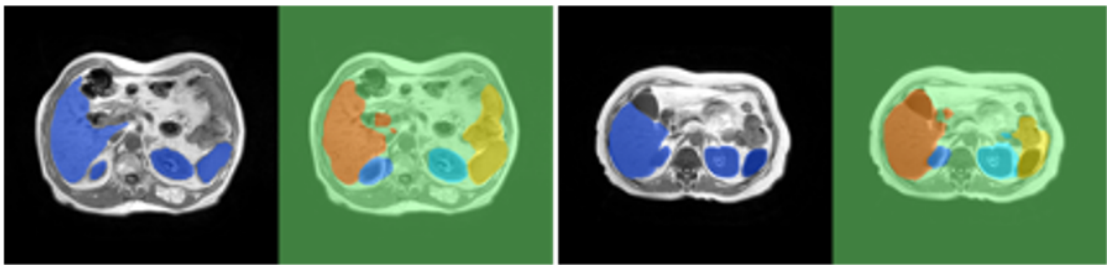

**Figure 2: Example slices with groundtruth and segmentation output (DeepLabV3).**

to its role in backpropagation learning. We reviewed the problems with metrics and formulated alternative loss functions, including different weighting on false positives and negatives. We created an experimental bench setup to compare and conclude regarding the various loss functions and alternatives. We concluded that IoU and dice and good choices, and that it can be worth experimenting with different weights. Future work on this issue should generalize to other datasets and problems and try to improve the formulation of loss function itself.

## ACKNOWLEDGMENTS

We acknowledge the organizers of the datasets [26] used in these experiments for the availability.

## REFERENCES

[1] Chen L, Bentley P, Rueckert D. Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks. NeuroImage Clin 2017; 15:633-43.
[2] Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, BengioY, *et al.* Brain tumor segmentation with deep neural networks. MedImage Anal 2017; 35:18-31.
[3] Choi H, Jin KH. Fast and robust segmentation of the striatum using deep convolutional neural networks. J Neurosci Methods2016; 274:146-53.
[4] Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. Med Phys2017; 44:547-57.
[5] Kline TL, Korfiatis P, Edwards ME, Blais JD, Czerwiec FS, Harris PC, *et al.* Performance of an artificial multi-observer deep neural network for fully automated segmentation of polycystic kidneys. J DigitImaging 2017; 30:442-8.
[6] Guo Y, Gao Y, Shen D. Deformable MR prostate segmentation via deep feature learning and sparse patch matching. IEEE Trans MedImaging 2016;35:1077-89.
[7] Li X, Dou Q, Chen H, Fu C-W, Qi X, Belav DL, *et al.* 3D multi-scale FCN with random modality voxel dropout learning for intervertebral disc localization and segmentation from multi-modality MR images.Med Image Anal 2018; 45:41-54.
[8] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, GhafoorianM, *et al.* A survey on deep learning in medical image analysis. MedImage Anal 2017; 42:60-88.
[9] Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin A, Do BT, WayGP, *et al.* Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface 2018;15.
[10] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).
[11] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.
[12] Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2017). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence, 40(4), 834-848.
[13] Zhou, X., Takayama, R., Wang, S., Zhou, X., Hara, T. and Fujita, H. (2017). Automated segmentation of 3D anatomical structures on CT images by using a deep convolutional network based on end-to-end learning approach. In Medical imaging 2017: image processing (Vol. 10133, p. 1013324). International Society for Optics and Photonics.
[14] Bobo M., Bao S., Huo Y., Yao Y., Virostko J., Plassard A. and Landman B. (2018). Fully convolutional neural networks improve abdominal organ segmentation. In Medical Imaging 2018: Image Processing (Vol. 10574, p. 105742V). International Society for Optics and Photonics.
[15] Larsson, M., Zhang, Y. and Kahl, F. (2016). Deepseg: Abdominal organ segmentation using deep convolutional neural networks. In Swedish Symposium on Image Analysis 2016.
[16] Chen, Y., Ruan, D., Xiao, J., Wang, L., Sun, B., Saouaf, R., Yang W., Li D. and Fan, Z. (2019). Fully Automated Multi-Organ Segmentation in Abdominal Magnetic Resonance Imaging with Deep Neural Networks. arXiv preprint arXiv:1912.11000.
[17] Groza, V., Brosch, T., Eschweiler D., Schulz, H., Renisch, S. and Nickisch, H. (2018). "Comparison of deep learning-based techniques for organ segmentation in abdominal CT images," in 1st Conference on Medical Imaging with Deep Learning (MIDL 2018), Amsterdam, The Netherlands, pp. 1-3, 2018. ,Úí pages 15, 16.
[18] Conze, P., Kavur, A., Gall, E., Gezer, N., Meur, Y., Selver, M. and Rousseau, F. (2020). Abdominal multi-organ segmentation with cascaded convolutional and adversarial deep networks. arXiv preprint arXiv:2001.09521.
[19] Fu Y., Mazur, T., Wu, X., Liu, S., Chang, X., Lu, Y., Harold, H., Kim, H., Roach, M., Henke, L. and Yang, D. (2018). A novel MRI segmentation method using CNN-based correction network for MRI-guided adaptive radiotherapy. Medical physics, 2018. 45(11): p. 5129-5137.
[20] Chlebus, G., Meine, H., Thoduka, S., Abolmaali, N., van Ginneken, B., Hahn, H., and Schenk, A. (2019). Reducing inter-observer variability and interaction time of MR liver volumetry by combining automatic CNN-based liver segmentation and manual corrections. PloS one, 14(5), e0217228.
[21] Hu, P., Wu, F., Peng, J., Bao, Y., Chen, F. and Kong, D. (2017). Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets. International journal of computer assisted radiology and surgery 12(3) (2017) 399–411.
[22] Wang Y., Zhou Y., Shen W., Park S., Fishman E. and Yuille A. (2019). Abdominal multi-organ segmentation with organ-attention networks and statistical fusion. Medical image analysis, 2019. 55: p. 88-102.

[23] Roth, R., Shen C., Oda H., Sugino T., Oda M., Hayashi H., Misawa K. and Mori K. (2018). A multi-scale pyramid of 3D fully convolutional networks for abdominal multi-organ segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention, 2018, pp. 417–425.

[24] Cai, J., Lu, L., Zhang, Z., Xing, F., Yang, L., Yin, Q.: Pancreas segmentation in MRI using graph-based decision fusion on convolutional neural networks. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MIC- CAI 2016. LNCS, vol. 9901, pp. 442-450. Springer, Cham (2016). doi:10.1007/ 978-3-319-46723-8 51.

[25] Zhang, X., Thibault, G., Decenciere, E., Marcotegui, B., Laø, B., Danno, R. & Chabouis, A. *et al.* (2014). Exudate detection in color retinal images for mass screening of diabetic retinopathy. Medical image analysis, 18(7), 1026-1043.

[26] Kavur A., Sinem N., Bar M., Conze P., Groza V., Pham D.,Chatterjee S., Ernst P., Ozkan S., Baydar B., Lachinov D., Han S., Pauli J., Isensee F., Perkonigg M., Sathish R., Rajan R., Aslan S., Sheet D., Dovletov G., Speck O., Nurnberger [27] A., Maier-Hein K., Akar B.,Unal G., Dicle O. and Selver M, (2020). CHAOS Challenge - Combined (CT-MR) Healthy Abdominal Organ Segmentation. In arXiv preprint, Jan. 2020. https://arxiv.org/abs/2001.06535CHAOS data a DOI number: https://doi.org/10.5281/zenodo.3362845.