ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

Cost-free resolution enhancement in Convolutional Neural Networks for medical image segmentation

Oscar J. Pellicer-Valero[1], María J. Rupérez-Moreno[2]
and José D. Martín-Guerrero[1]

[1] Intelligent Data Analysis Laboratory, Department of Electronic Engineering,
ETSE (Engineering School), Universitat de València (UV), Spain.
[2] Centro de Investigación en Ingeniería Mecánica (CIIM),
Universitat Politècnica de València (UPV), Spain.
E-Mail: Oscar.Pellicer@uv.es, mjrupere@upvnet.upv.es, jose.d.martin@uv.es

Abstract. High-resolution segmentations of medical images are imperative for applications such as treatment planning, image fusion or computer-aided surgery. Nevertheless, these are often hard and time-consuming to produce. This paper presents a method for improving the output resolution of Convolutional Neural Networks (CNNs) for medical image segmentation. It is straightforward to implement and works with any already trained CNN with no modification nor retraining required. It is able to produce better results than binary interpolation methods since it exploits all the contextual information to predict the sought values.

1 Introduction

Segmentation in medical images is a voxel-level classification task such that all voxels corresponding to a particular class represent a single semantical entity in the body: an organ, a bone, a tissue, a lesion, etc. Segmentation algorithms take an image as an input (e.g., a chest radiography), and one or several masks (e.g., lungs and lesions) are obtained as an output (Figure 1). These algorithms are commonly applied to medical imaging techniques such as Magnetic Resonance (MR), Computerized Tomography (CT) and Ultrasound (US).

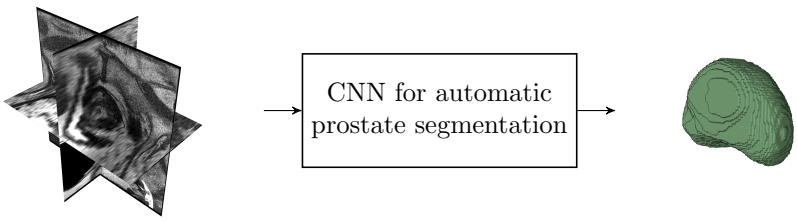

Fig. 1: A CNN trained on the task of prostate segmentation takes a 3D prostate MR image as an input and obtains a 3D prostate mask as an output.

Obtaining accurate segmentations is a very valuable yet difficult endeavor. On one hand, segmentations are valuable as they are mandatory inputs for

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

image-based diagnosis, lesion detection and treatment planning; furthermore, for three-dimensional (3D) images, the obtained geometries can be used to perform simulations of the biomechanical behavior of a body, which can then be used in image fusion, surgical planning, computer-aided surgery or bone-strength simulations, to cite a few applications.

On the other hand though, accurate segmentations are hard and laborious to obtain, since they have to be manually annotated by expert radiologists, and even then, the inter- and intra-observer variations may be significant [1]. Because of this, automatic segmentation algorithms for medical images have become increasingly prevailing.

Although several different automatic segmentation frameworks have been suggested in the past, current state-of-the-art techniques usually employ Convolutional Neural Networks (CNNs) and, more specifically, those based on the U-Net architecture [2], which has lead to segmentation accuracies above the inter-observer threshold in increasingly more scenarios [1].

High-resolution segmentations are often essential in the aforementioned applications. However, automatic segmentation techniques tend to present two closely related problems. Firstly, CNNs usually require the input image to be downscaled before processing it to alleviate the Graphics Processing Unit (GPU) memory costs associated with 3D convolutions. Secondly, so-called 3D medical imaging techniques are often actually two-dimensional (2D) multi-slice images instead, which are then stacked to form the final 3D geometry. These images, however, usually have a dimension (perpendicular to all the individual 2D slices) along which the resolution is much coarser than the rest. For instance, the MR image in Figure 1 suffers from this inconvenience.

Even if the first problem could be solved by using a finer resolution image as input to the CNN, the second problem remains still a challenge, since no inter-slice information can be extracted from 2D multi-slice images in order to perform a finer segmentation.

One possible solution would be to improve the resolution of the input image in an intelligent manner, which is a problem known as super-resolution [3]. These upscaled images could then be used to train a segmentation CNN, thus obtaining higher resolution output masks. Some works have already studied this in the medical domain; for instance, [4] proposes a CNN to upscale 2D multi-slice images of the heart along the axis perpendicular to the slices, achieving perceptible improvements. However, with this approach, the problem of GPU memory limitations still remains. Furthermore, in order to train the CNN, many images should be manually segmented at the new increased resolution, thus making the process even more time-consuming.

Another possible solution is to employ binary interpolation techniques to produce a high-resolution mask from a lower resolution one. The simplest approach is nearest-neighbor interpolation, which simply takes the value of the closest neighbor for any given point. Even if this procedure produces very "blocky" low-quality interpolations, it is still widely used due to its simplicity and speed. A better approach consists in taking any kind of interpolator for real numbers

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

and using it to interpolate the binary masks for each class independently; then, the class with the highest value at any given point is used as the final label for that point. This approach provides much smoother results and, in combination with linear interpolation, it is also very quick. Finally, some more complex algorithms have been proposed to deal specifically with the problem of inter-slice interpolation in 2D multi-slice images, such as in [5]. However, all these methods present one important pitfall: they completely disregard the contextual information contained either in the original image or in the domain knowledge of the problem. The single notable exception seems to be [6], where the authors combine both the binary morphology and the local intensities to perform the interpolation.

In this paper, a method for intelligent upscaling of the output mask of a medical image segmentation CNN is proposed. This method takes into account all the available contextual information and it is cost-free in the sense that it can be applied to any already trained CNN with no modifications to its architecture or any retraining required.

2 Materials and methods

The proposed method exploits a very simple yet effective idea for performing intelligent output upscaling on already trained segmentation CNNs. It consists in shifting the input image by several different sub-voxel amounts, feeding these transformed images to the CNN in order to obtain the segmentation masks, and then combining them into a single final high resolution mask. Despite its simplicity, this procedure achieves high resolution segmentation masks which outperform other discussed approaches (as it will be discussed in Section 3) from already trained (an possibly low resolution) segmentation CNNs. Thus, the problem of interpolation is shifted from the mask domain to the image domain, where the conveyed information is still complete and not yet binarized.

For a more detailed description of the method, consider a CNN with input and output dimensionality (or resolution) of $(d_1 \times \cdots \times d_N)$, where N is the number of dimensions (e.g., $N = 3$ for a 3D image). Suppose we wanted to increase the output resolution along a single dimension i by an integer factor of k_i, such that the output resolution were: $(d_1 \times \cdots \times d_i \cdot k_i \times \cdots \times d_N)$.

First, we would need to generate $k_i - 1$ images $[I_1, \ldots, I_{k_i-1}]$ from the original input image I_0, each one shifted $+\frac{1}{k_i}$ voxels along dimension i with respect to the previous one. Therefore, in order to obtain $[I_1, \ldots, I_{k_i-1}]$, I_0 must be interpolated and evaluated at the positions given by the translation transforms $\left[(0, \ldots, \frac{1}{k_i}, \ldots, 0), \ldots (0, \ldots, \frac{k_i-1}{k_i}, \ldots, 0)\right]$ applied to I_0, where the translation has a value of zero for all dimensions except for i. Second, all the k_i images $[I_0] \cup [I_1, \ldots, I_{k_i-1}]$ are fed to the CNN one by one, and k_i outputs masks $[O_0, O_1, \ldots, O_{k_i-1}]$ are obtained in return. Finally, $[O_0, O_1, \ldots, O_{k_i-1}]$ are combined by interleaving them voxel-wise along i in order to obtain a single output mask $O_{combined}$, which will have a k_i-times higher resolution along axis i. In this context, interleaving can be defined as stacking $[O_0, O_1, \ldots, O_{k_i-1}]$ to pro-

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

duce $O_{combined}$ in such a way that the n^{th} slice along dimension i in $O_{combined}$ corresponds to the $\lfloor \frac{n}{k_i} \rfloor^{th}$ slice along dimension i of $O_{n\%k_i}$. Figure 2 provides a visual representation of the described methodology.

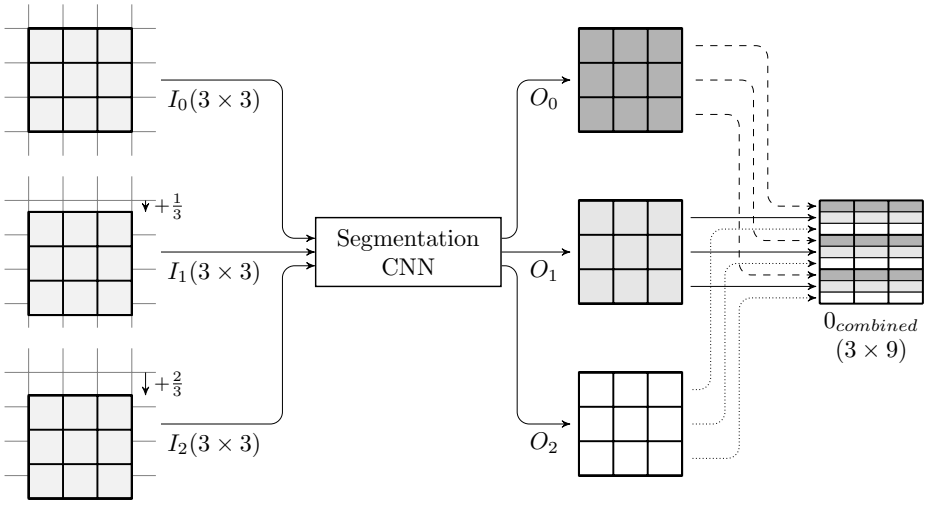

Fig. 2: Visual representation of the method for an image of dimensions $(d_1, d_2) = (3, 3)$, for $k_2 = 3$.

The proposed method can be extended to simultaneously improve the resolution of the output mask along any number of dimensions. As an overview, a new set of transformations T must be computed and applied to I_0, by combining in all possible ways the transformations which would be required to increase the resolution by a factor k_m along a single dimension $m \in (1, N)$ independently:

$$T = \left[(\frac{1}{k_1}, 0, \dots, 0), \dots (\frac{k_1 - 1}{k_1}, 0, \dots, 0) \right] \times \cdots \\ \times \left[(0, \dots, 0, \frac{1}{k_N}), \dots (0, \dots, 0, \frac{k_N - 1}{k_N}) \right] \quad (1)$$

Finally, the resulting masks are combined to form the final mask. It must be however noted that, for some positions in $O_{combined}$ there will be several possible values due to overlap among masks. In those instances, a binary fusion function (such as the majority vote) must be used to produce a final value.

The computational cost of the method is approximately $c_0 \cdot \prod_{i=1}^{N} k_i$, where c_0 is the cost of interpolating an image and passing it through the CNN. This cost comes from computing the size of the set of transformations shown in Eq. (1). As an example, if we wanted to increase the resolution along a single axis i by a factor of k_i, the cost would be k_i times the cost of obtaining a single mask in the native CNN resolution.

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2.
Available from http://www.i6doc.com/en/.

3 Results and discussion

This method was applied to a segmentation CNN trained on several datasets [7, 8] of prostate MR 2D multi-slice images, where the approximate physical voxel spacings are $(s_x, s_y, s_z) \approx (0.5, 0.5, 3)mm$, and s_z is to be improved by a factor $k_z = 3$ ($s_z = 1mm$) and $k_z = 6$ ($s_z = 0.5mm$). Figure 3 shows a comparison between two of the most common binary interpolation methods and the proposed method.

The proposed method produces the smoothest results of the three, as it can be noticed by comparing the upper slices of the prostate masks in Figure 3. Furthermore, it does not just interpolate between slices, but rather predicts the mask at several inter-slice levels. Therefore, it is able to obtain more accurate results by incorporating the underlying image as input, as well as all the contextual information that the CNN has learned about the problem of segmenting a particular part of the body. Unfortunately, no numerical results can be provided, as no ground truth is available, since all the methods are interpolating beyond the resolution of the original image.

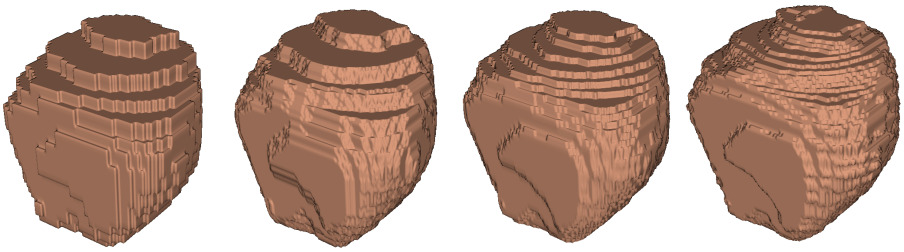

Fig. 3: Mask predicted by a prostate segmentation CNN and upscaled along the z-axis using (from left to right): nearest-neighbor interpolation, Gaussian interpolation ($k_z = 6$), the proposed method with $k_z = 3$ and the proposed method with $k_z = 6$.

This approach seems to be closely related to a technique known as Test-Time Augmentation, wherein an already trained CNN is provided with several randomly augmented (translated, rotated, shifted, etc.) versions of the same input image, and the outputs are combined into a single output prediction, which is oftentimes more accurate than any individual prediction. Similarly, the proposed method feeds the CNN several transformed versions of the same input and then combines all the outputs. However, by contrast, the transformations are not random and follow instead a very precise structure which must also be taken into account in the output combination process.

4 Conclusion and further work

This paper presents a method for intelligently improving the output resolution of CNNs for medical image segmentation. It is better than other uspcaling methods

ESANN 2020 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Online event, 2-4 October 2020, i6doc.com publ., ISBN 978-2-87587-074-2. Available from http://www.i6doc.com/en/.

since it does not perform interpolation, but rather it predicts the sub-voxel values using the image and the context information that the CNN has encoded about the particular problem. It can be used to improve the resolution of 2D multi-slice images beyond the original resolution of the image, thus providing an accurate 3D segmentation for methods that require it, such as in the simulation of the biomechanical behavior of a body. Finally, it is a very simple to implement post-processing step that can make use of any already existing CNN with no modifications required whatsoever.

As a main downside, the method can only improve the resolution of the predictions of a CNN, unlike general binary interpolation algorithms, which can upscale any binary image. Also, the computational cost of this procedure can be high if the resolution is increased along many different axes simultaneously. Lastly, it is not proven in any way that the results should be smooth and/or correct, and it is instead trusted in the empirical results and the robustness intuitions about CNN architectures.

From the ideas here presented, two main research lines arise. First, it should be explored how well this technique generalizes to natural image segmentation CNNs, and how useful it would be in this context. Second, and more interestingly, a niche for improvement has been discovered in binary interpolation algorithms for segmentations. Namely, almost all current binary interpolators disregard the precious information contained in the original image to perform the interpolation. A clever exploitation of this information could yield improved interpolations for segmentation masks.

References

[1] Bulat Ibragimov and Lei Xing. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med. Phys.*, 44(2):547–557, feb 2017.

[2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, volume 9351, pages 234–241. Springer Verlag, 2015.

[3] Christian Ledig, Lucas Theis, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2017*, volume 2017-Janua, pages 105–114, 2017.

[4] Ozan Oktay, Wenjia Bai, et al. Multi-input cardiac image super-resolution using convolutional neural networks. In *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, volume 9902 LNCS, pages 246–254, 2016.

[5] Alexandra Branzan Albu, Trevor Beugeling, and Denis Laurendeau. A morphology-based approach for interslice interpolation of anatomical slices from volumetric images. *IEEE Trans. Biomed. Eng.*, 55(8):2022–2038, aug 2008.

[6] Xiaochun Liao, David Reutens, and Zhengyi Yang. Morphology-based interslice interpolation using local intensity information for segmentation. In *Proc. - 2011 4th Int. Conf. Biomed. Eng. Informatics, BMEI 2011*, volume 1, pages 384–389, 2011.

[7] Geert Litjens, Robert Toth, et al. Evaluation of prostate segmentation algorithms for MRI: The PROMISE12 challenge. *Med. Image Anal.*, 18(2):359–373, feb 2014.

[8] Geert Litjens, Futterer Jurgen, and Henkjan Huisman. Data From Prostate-3T, 2015.