# Can multiple segmentation methods enhance deep learning networks generalization? A novel hybrid learning paradigm

PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

# Can multiple segmentation methods enhance deep learning networks generalization? A novel hybrid learning paradigm

Francesco Marzola, Kristen Meiburger, Filippo Molinari, Massimo Salvi

Francesco Marzola, Kristen M. Meiburger, Filippo Molinari, Massimo Salvi, "Can multiple segmentation methods enhance deep learning networks generalization? A novel hybrid learning paradigm," Proc. SPIE 12465, Medical Imaging 2023: Computer-Aided Diagnosis, 1246516 (7 April 2023); doi: 10.1117/12.2653394

SPIE.

Event: SPIE Medical Imaging, 2023, San Diego, California, United States

# Can multiple segmentation methods enhance deep learning networks generalization? A novel hybrid learning paradigm

Francesco Marzola[a], Kristen M. Meiburger*[a], Filippo Molinari[a], Massimo Salvi[a]
[a] Biolab, Department of Electronics and Communications, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Torino, Italy

**ABSTRACT**

Deep learning methods are the state-of-the-art for medical imaging segmentation tasks. Still, numerous segmentation algorithms based on heuristic-based methods have been proposed with exceptional results. To validate segmentation algorithms, manual annotations are typically considered as ground truth. However, manual annotations often suffer from inter/intra-operator variability and can also be occasionally inaccurate, especially when considering time-consuming and precise tasks. A sample case is the manual delineation of the lumen-intima (LI) and media-adventitia (MA) borders for intima-media thickness (IMT) measurement in B-mode ultrasound images.
In this work, a novel hybrid learning paradigm which combines manual segmentations with the automatic segmentation of a dynamic programming technique for ground truth determination is presented. A profile consensus strategy is proposed to construct the hybrid ground truth.
Two open-source datasets (n=2576) were employed for training four deep learning networks using the hybrid learning paradigm and three single source training targets as a comparison. The pipeline was fixed across the four tests and included a Faster R-CNN detection network to locate the carotid artery and then subsequent division into patches which were segmented using a UNet. The validation of the results was performed on an external test set comparing the predictions of the four different models to the annotations of three independent manual operators.
The hybrid learning paradigm showed the best overall segmentation results (Dice=0.907±0.037, $p<0.001$) and demonstrated an exceptional correlation between the mean of three operators and the automatic measure (ICC(2,1)=0.958), demonstrating how the incorporation of heuristic-based segmentation methods within the learning paradigm of a deep neural network can enhance and improve final segmentation performance results.

**Keywords:** Deep learning, segmentation, hybrid ground truth, intima-media thickness, ultrasound, controllable AI, UNet

## 1. INTRODUCTION

In the last decade, deep learning-based methods have risen in popularity for computer vision tasks for the analysis of both natural images and for medical imaging. These methods tend to have a clear performance advantage when compared to heuristic-based methods, as they are able to better generalize the task, but have the drawback of needing a large amount of properly annotated data for model training. The annotated data are typically referred to as "ground truth", but are often far from being perfect, as they are usually obtained manually and are hence dependent on the operator experience, concentration, and precision. For numerous medical imaging tasks, there is a vast literature of heuristic-based automated algorithms that have shown good performance results. One such example is the segmentation of the Intima Media Complex (IMC) in B-mode ultrasound images, that is often based on the extraction of the Lumen-Intima (LI) and Media-Adventitia (MA) boundaries. The distance between the determined LI and MA borders is computed and taken as the Intima-Media Thickness (IMT), a marker for atherosclerosis risk assessment. In this context, numerous deep learning-based approaches have been proposed, typically making use of segmentation networks like UNet and its variants to achieve automatic or semi-automatic tracing of the LI and MA boundaries. Some of these methods have already reached performance in par or lower than inter-operator variability [1-5]. Still, the current analyses often lack a real out-of-distribution test set and rely on cross validation techniques to assess the generalization ability of the models. Moreover, the ground truth that serves as

*kristen.meiburger@polito.it; phone +39 011 090 4109; biolab.polito.it

Medical Imaging 2023: Computer-Aided Diagnosis, edited by Khan M. Iftekharuddin, Weijie Chen, Proc. of SPIE Vol. 12465, 1246516 · © 2023 SPIE
1605-7422 · doi: 10.1117/12.2653394

a training target for the networks, is typically derived from the manual segmentation of a single operator, thus intrinsically biasing the network predictions towards one specific operator. As automatic IMT measurement has been widely studied in literature before deep learning approaches took the place of heuristic-based models [6], these studies should not be ignored as they can be a viable tool to enhance the performance of deep learning systems, making them even more generalizable.

Here we propose a novel hybrid learning paradigm that exploits the best of both the heuristic and the deep learning worlds in the specific case of IMT measurement, combining manual segmentations with the robust automatic segmentation of a dynamic programming technique [7]. The collectively determined segmentation is then employed as the annotated ground truth for model training and validation.

The main contributions of this work are the following:

- A novel approach of employing a robust traditional image processing method as a viable way to fine tune manual segmentations, providing a hybrid and collective ground truth annotation for training deep learning models.
- A quantitative comparison proving how the innovative hybrid ground truth learning paradigm enables deep learning models to further generalize the learning process, thus improving the performance when tested on an out-of-distribution test set.

## 2. MATERIALS AND METHODS

### 2.1 Dataset Description

Two previously published and freely downloadable datasets [6],[8], were used as the training and validation sets. The resulting dataset consisted of 2576 B-mode longitudinal ultrasound images of the carotid artery. Acquisition details are available in the original papers. Two manual annotations of the LI and MA profiles from the same expert analyst at two time points (A1 and A1s) and one computerized measurement based on dynamic programming [7] were employed for the hybrid learning paradigm, and are also freely downloadable [6],[8]. The dynamic programming computerized method was developed by researchers of the Technische Universität München and for simplicity is referred to here as TUM.

To test the proposed method, an external dataset consisting of 465 images from 4 different centers was used, for which three manual tracings were available [9][10].

### 2.2 Hybrid learning paradigm and profile consensus strategy

For the hybrid learning paradigm, a profile consensus strategy was proposed to create a collectively determined hybrid ground truth (Hybrid GT) profile by performing a column-wise consensus separately for the LI and MA borders. First, a similarity parameter was employed to measure the consistency between two profiles ($P_1$ and $P_2$, respectively) on their common support, determined as the columns of the image where both profiles are defined:

$$Similarity\ (P_1;P_2) = \frac{Correlation\ (P_1;P_2)}{Bias\ (P_1;P_2)} \tag{1}$$

*Correlation*($P_1$; $P_2$) is calculated between the row values of the two LI/MA profiles, while *Bias*($P_1$; $P_2$) is the mean column-wise difference between the two profiles. The range for similarity values for the LI profile was found to be [0, 3.30] and [0, 3.82] for the MA profile. Profiles were determined as being *similar* if the computed Similarity was over the threshold of 0.5, determined empirically. As two manual and one computerized segmentation (i.e., A1, A1s, and TUM, respectively)

were considered, the Similarity parameter was hence computed for the three different cases: A1 vs. A1s, A1 vs. TUM, and A1s vs. TUM.

For the Hybrid GT profile determination, five conditions can occur, which are shown in Figure 1:

1. No manual or computerized profile point is present: in this case, no Hybrid GT profile point is defined.
2. One profile is present: that profile defines the Hybrid GT.
3. Two manual profiles are present: Hybrid GT is the mean between the two profiles if they present a high similarity; if not, only the most similar to the computerized method is kept.
4. Two profiles are present, one manual and one computerized: Hybrid GT is the mean between the two profiles if they present a high similarity; if not, only the manual profile is kept.
5. All three profiles are present: the mean between the two profiles with the highest similarity between the three determines the Hybrid GT profile point.

An illustration of this process is available in Figure 1, where the different choices between LI and MA profiles can be seen. In Figure 1, the two manual LI profiles are similar and averaged to obtain the Hybrid GT. In contrast, the two manual MA profiles have a low similarity, hence the TUM profile is averaged with the A1 profile, as it presented the highest similarity to the TUM profile.

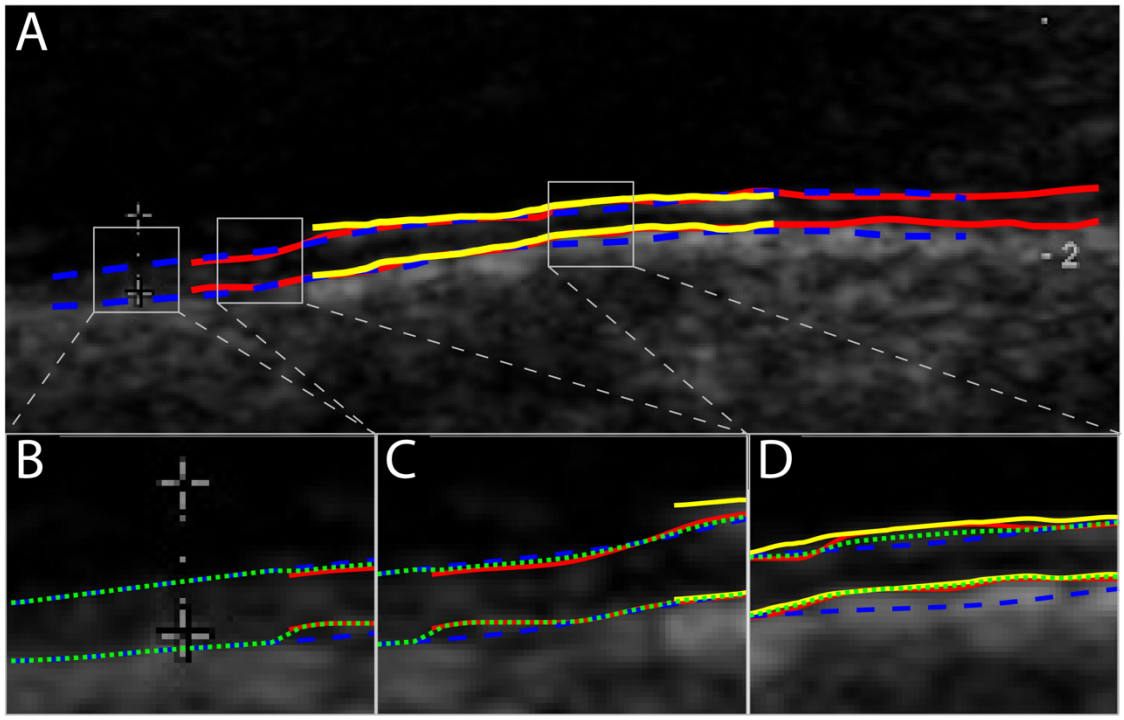


Figure 1. Illustration of the Hybrid GT construction. The profiles are A1 (red), A1s (blue dashed), TUM (yellow), Hybrid GT (green dotted). In panel A there are the three available profiles (two manual, A1 and A1s, and one computerized, TUM). In panel B, the condition where only one profile is available is shown. In panel C, two profiles are available. In panel D, three profiles are again available.

### 2.3 Deep learning detection and segmentation models – training and inference

The developed deep learning system consisted of two distinct parts. The first one had the goal of detecting the Region Of Interest (ROI) that encloses the IMC by applying a detection network. The second part used a patch-based segmentation network to obtain a segmentation mask of the IMC (Figure 2).

For training both the detection and segmentation deep learning models, the dataset was split with a 90/10 ratio resulting in 2311 images in the training set and 265 in the validation set. The original image was first cropped using a heuristic-based method described in [8] to include only the useful ultrasound image, and then the cropped image was resized to 480x480 pixels. The same transformations were also applied to the mask created using the Hybrid GT profile.

A Faster R-CNN was applied as a detection network [11] to extract the area in which to segment the LI and MA borders. The Faster R-CNN is a region proposal network that simultaneously predicts object boundaries and scores for each detected bounding box. This detection network consists of two modules: the first module is a deep convolutional network (ResNet50) that proposes regions, and the second module is a detector [12] which analyzes and classifies the proposed regions. The proposed detection network employs anchor boxes that consist of a reference box with a specific scale and aspect ratio. Specifically, we used a total of five reference anchor boxes ([150, 170, 190, 220, 240]) with three aspect ratios ([2, 3.5, 5]) to detect objects with various shapes. The network was trained for 20 epochs with an early stopping patience of 5 epochs when performance stopped increasing on the validation set. The learning rate was set to $10^{-3}$ with a batch size equal to 6. Next, up to six 96x96 patches from the output of the detection network were extracted with MONAI (Medical Open Network for AI) [13]. As a result, 13862 patches were included in the training set and 1590 patches in the validation set for the subsequent segmentation network.

For the segmentation task, a standard UNet with a Resnet50 pretrained encoder was implemented (batch size = 32, Adam optimization, LR 0.0002, Dice loss). Online data augmentation was applied on the patches selected for training with a random permutation of the following transformations: Scaling and Rotation (scale = ± 10%, rotate = ±10°, p = 1), Horizontal flipping (p = 0.5), Gaussian Noise Addition (p = 0.3), Multiplicative Noise Addition (p = 0.2), Contrast Limited Adaptive Histogram equalization (CLAHE) (p = 0.3), Sharpening (p = 0.25), Optical Distortion (p = 0.2), and Blurring (p = 0.5). The best model was chosen as the one yielding the best Intersection over Union on the validation set. Our approach was developed using Pytorch and with the detectron2, Albumentations and Segmentation Models libraries [14 - 17]. Training was performed on an RTX 3090 with 24 GB of VRAM. Training time was 1.5 hours for the detection model and 0.5 hours for the segmentation model.

During inference on the external test set, the same pipeline of detection followed by patch extraction and segmentation was applied. The network prediction was then padded and upsampled to the original image dimensions to extract the LI and MA profiles with the same reference system of the manual profiles. Then, the boundary of the biggest connected area was extracted and a moving average with a window of 5 pixels was applied to the coordinates extracted from the perimeter, to smoothen the final segmented profile.

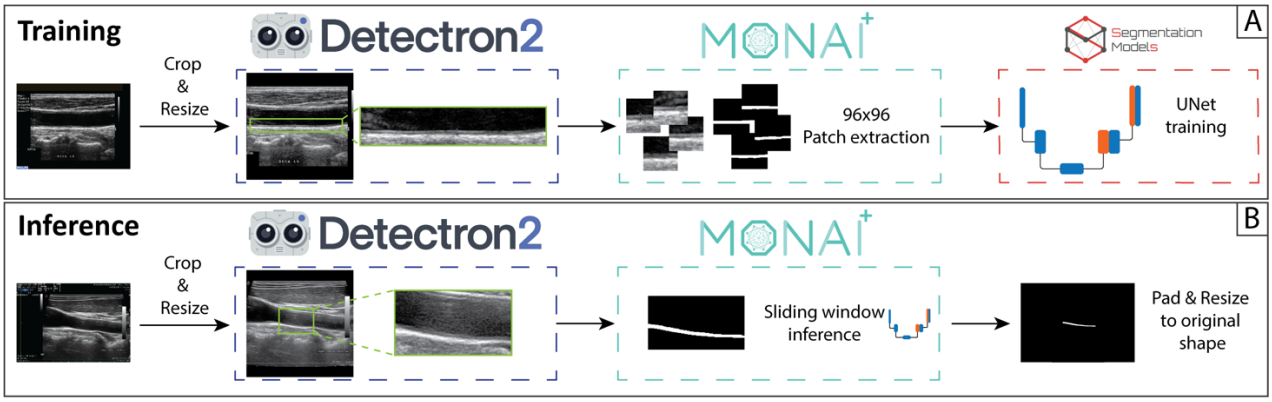


Figure 2. Training (top row) pipeline and inference (bottom row) pipeline.

### 2.4 Validation

To determine which learning paradigm (that is, when using the manual, computerized or Hybrid GT profile as the ground truth target for training the networks) gave the best results on the test set, the same network was trained with 4 different targets (i.e., A1, A1s, TUM, and Hybrid GT) and inference was then performed on the same external test set. The resulting predictions from the UNet were named according to the training target (i.e., $UNet_{A1}$, $UNet_{A1s}$, $UNet_{TUM,}$ $UNet_{HybridGT}$).

The reference mask and the reference LI and MA profiles for the test set were obtained from the consensus of three independent manual tracings (GT1, GT2, GT3, respectively). The consensus was built using the same approach previously described for the hybrid ground truth creation.

To assess the performance of the segmentation framework, the Dice coefficient and the Hausdorff distance (HD) were computed between the predicted and the reference masks. To specifically assess the precision on LI and MA boundaries, the Hausdorff distance (HD) was also measured between the predicted and reference LI and MA profiles. Furthermore, the absolute IMT bias was computed, defined as:

$$Abs.\ IMT\ Bias = /IMT_{method} - IMT_{Ground\ truth}/ \quad (2)$$

where *method* refers to the automatic segmentation and *Ground truth* is the IMT measured from the consensus of the three manual operators. The validation parameters were computed considering the common support (i.e., the columns where both the target and all the predicted profiles are defined) and are shown in Table 1. For each metric, to assess the statistical significance of the differences between the best performing method and the others, a Wilcoxon test was performed, and the non-normality of the distribution was checked with a chi2 test.

To further confirm the generalization performance on the test set, the IMT measurements from the manual operators and from the hybrid learning paradigm were compared. The intraclass correlation coefficient (ICC(2,1)) was computed considering two cases: (1) between the IMT values obtained by the three manual operators and (2) between the IMT value obtained by $UNet_{HybridGT}$ and the average IMT value between the three operators.

To assess the $UNet_{HybridGT}$ agreement with manual operators compared to inter-operator variability, a Bland-Altman analysis was performed.

## 3. RESULTS

The proposed Hybrid GT showed the best segmentation metrics with a Dice score of 0.902 ± 0.047 and a HD between the segmentation masks of 3.53 ± 3.05 mm. The analyses showed better performances for the $UNet_{HybridGT}$ for the tracing of the LI profile with a Hausdorff Distance of 0.158 mm ± 0.092 mm. On the other hand, the network trained only with the TUM ground truth outperformed the novel approach for the IMT absolute bias (0.053 mm ± 0.088 mm vs. 0.136 ± 0.079 mm for the $UNet_{HybridGT}$) and the MA profile tracing (0.232 mm ± 0.208 mm vs. 0.272 ± 0.204 mm for the $UNet_{HybridGT}$).

The new approach's robustness was confirmed by the correlation analysis between manual and automatic measurements. The Pearson's correlation of IMT values between the $UNet_{HybridGT}$ and the three operators were 0.944, 0.889 and 0.944 for GT1, GT2 and GT3, respectively. Correlations between manual operators were 0.863 (GT1 vs. GT2), 0.925 (GT1 vs. GT3) and 0.827 (GT2 vs. GT3).

The analysis of the agreement between the measurements of the proposed method $UNet_{HybridGT}$ to the average of the manual measurements showed an excellent agreement with an ICC(2,1) of 0.958. This result was well above the agreement between the three manual operators that was found to be equal to 0.859.

The Bland-Altman analysis in Figure 3 showed a slight overestimation of the IMT value by $UNet_{HybridGT}$ with no signs of bias in the error distribution. The reproducibility coefficient (RPC) was similar when comparing the manual and automatic methods, thus confirming that the automatic method performs within the inter-operator variability.

| Table 1. Results summary – common support | | | | |
|---|---|---|---|---|
| | **$UNet_{A1}$** | **$Unet_{A1s}$** | **$Unet_{TUM}$** | **$Unet_{HybridGT}$** |
| **DICE** | 0.893 ± 0.043 | 0.877 ± 0.045 | 0.897 ± 0.051 | **0.907 ± 0.037*** |
| **HD mask (pixel)** | 3.53 ± 2.53 | 3.98 ± 2.23 | 3.75 ± 2.36 | **3.21 ± 2.14*** |
| **ABS IMT bias (mm)** | 0.153 ± 0.088 | 0.195 ± 0.107 | **0.053 ± 0.088*** | 0.136 ± 0.079 |
| **HD LI (mm)** | 0.176 ± 0.105 | 0.199 ± 0.140 | 0.199 ± 0.211 | **0.158 ± 0.092*** |
| **HD MA (mm)** | 0.301 ± 0.238 | 0.327 ± 0.226 | **0.232 ± 0.208*** | 0.272 ± 0.204 |

HD mask: Hausdorff distance between the reference mask and the predicted mask; ABS IMT bias: absolute intima-media thickness bias; HD LI(MA): Hausdorff distance between the reference LI(MA) profile and the predicted LI(MA) profile. $UNet_{A1}$, $UNet_{A1s}$, $UNet_{TUM}$, $UNet_{HybridGT}$: UNet models trained with the specified ground truth as the target. In bold the best performing method and the * indicates a significant difference with the other methods (Wilcoxon Test, $p \leq 0.01$).

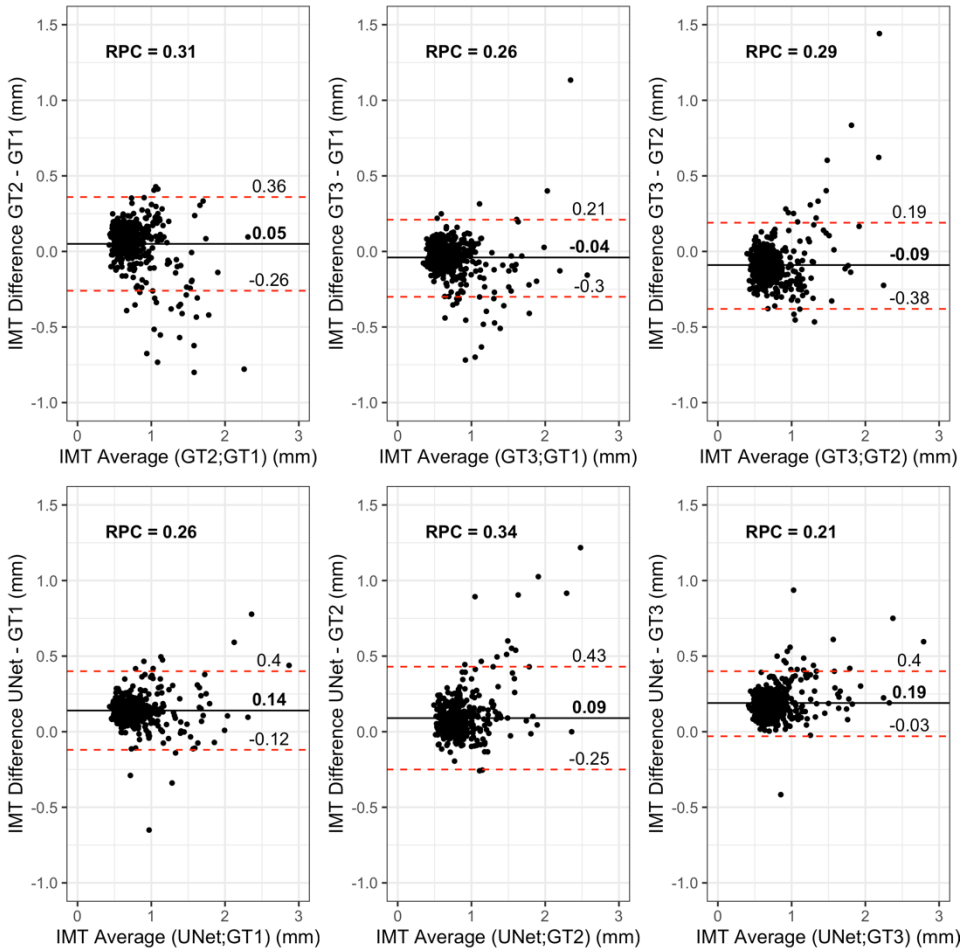


Figure 3. Bland-Altman plots comparing inter-operator variability and $UNet_{HybridGT}$ results on the test set. Manual measurements are referenced as GT1, GT2, GT3. $UNet_{HybridGT}$ is referenced as UNet. RPC: reproducibility coefficient (1.96*standard deviation of the errors). The horizontal lines define the mean error (black solid line), and upper and lower limits (dashed red lines, ± 1.96 SD). In bold the mean differences with a significant statistical difference ($p \leq 0.01$).

## 4. DISCUSSION

A novel hybrid learning paradigm which combines manual segmentations with the automatic segmentation of a dynamic programming technique for an innovative collective ground truth determination was presented, proving how computerized methods based on traditional segmentation techniques can enhance the generalization capability of deep learning networks. This was demonstrated by the fact that the network trained with the novel paradigm for the ground truth definition achieved better mask segmentation performances compared to the use of only manual or only computerized segmentation as the ground truth used for model training and validation.

The measurement of the Hausdorff Distance between the profiles obtained by $UNet_{HybridGT}$ with those obtained through the consensus strategy on the three manual profiles showed mixed results. $UNet_{HybridGT}$ achieved better results for LI while being less precise compared to $UNet_{TUM}$ in MA tracing and IMT measurement. Still, $Unet_{HybridGT}$ performed better than $Unet_{A1}$ and $Unet_{A1s}$ and had a lower standard deviation with respect to $Unet_{TUM}$. proving its robustness. Moreover, in the test set the mean conversion factor was equal to $0.0618 \pm 0.0109$ mm/pixel, hence the difference for HD MA between $UNet_{TUM}$ ($0.232 \pm 0.208$ mm) and $UNet_{HybridGT}$ ($0.272 \pm 0.204$ mm) lies in less than one pixel.

One possible explanation of this behavior is that the MA boundary is less affected by the noise typical of ultrasound images [6][8], thus it can usually be traced with high accuracy by a heuristic-based computerized method. As a result, the network trained with only that information can transfer better this knowledge on an external test set. Conversely, LI tracing relies more on human experience, hence the networks that have been provided with hybrid computerized and manual examples can trace that boundary with greater accuracy.

Introducing in the training examples information derived from a computerized method helps the network learn more strict rules. This learning process minimizes the errors when segmenting targets well described by those rules as in the case of the MA boundary definition. This is shown also when training the network with a ground truth derived only from the computerized method. In this case, the network lost some generalization capability, but achieved better performance when considering finer details of the segmentation of the MA profile.

It is important to underline that, while shown here for the specific case of carotid artery IMT measurement, the proposed approach can be applied to a myriad of different tasks in automatic medical image analysis. While traditional heuristic-based methods alone lack performance compared to deep learning-based methods, they can be used in the training phase of deep learning models to enhance their generalization capability and mitigate errors due to inaccurate manually annotated data. Hence, all the knowledge from heuristic-based methods that has been gained throughout the years can prove to be a crucial stepping-stone to boost modern deep learning systems and achieve better results for specific tasks. This collective integration between manual and computerized labels can provide the networks with examples of labels based both on operator experience and on physics and mathematical principles, thus guiding the training of the network towards respecting both constraints.

## 5. CONCLUSION

In conclusion, the results shown here demonstrated how the proposed novel hybrid learning paradigm which incorporates not only manual segmentations but also a computerized method based on heuristic-based segmentation techniques can enhance the generalization capability of deep learning networks and can be a valued asset to increase their performance. This innovative hybrid approach can be applied to a plethora of medical imaging tasks where there are established heuristic algorithms – doing so will enable the advancement of current deep learning systems to learn fundamental ancillary rules from traditional image processing methods alongside the ones learned from manual annotations, bridging together the best of both approaches.

## REFERENCES


[1] N. Lainé, G. Zahnd, H. Liebgott and M. Orkisz, "Segmenting the carotid-artery wall in ultrasound image sequences with a dual-resolution U-net," 2022 IEEE International Ultrasonics Symposium (IUS), Venice, Italy, 2022, pp. 1-4, doi: 10.1109/IUS54386.2022.9957590.

[2] Mainak Biswas, Venkatanareshbabu Kuppili, Tadashi Araki, Damodar Reddy Edla, Elisa Cuadrado Godia, Luca Saba, Harman S. Suri, Tomaž Omerzu, John R. Laird, Narendra N. Khanna, Andrew Nicolaides, Jasjit S. Suri,

Deep learning strategy for accurate carotid intima-media thickness measurement: An ultrasound study on Japanese diabetic cohort, Computers in Biology and Medicine, Volume 98, 2018, Pages 100-117, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2018.05.014.

[3] Mainak Biswas, Luca Saba, Shubhro Chakrabartty, Narender N. Khanna, Hanjung Song, Harman S. Suri, Petros P. Sfikakis, Sophie Mavrogeni, Klaudija Viskovic, John R. Laird, Elisa Cuadrado-Godia, Andrew Nicolaides, Aditya Sharma, Vijay Viswanathan, Athanasios Protogerou, George Kitas, Gyan Pareek, Martin Miner, Jasjit S. Suri, Two-stage artificial intelligence model for jointly measurement of atherosclerotic wall thickness and plaque burden in carotid ultrasound: A screening tool for cardiovascular/stroke risk assessment, Computers in Biology and Medicine, Volume 123, 2020, 103847, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2020.103847.

[4] Yanchao Yuan, Cancheng Li, Lu Xu, Shangming Zhu, Yang Hua, Jicong Zhang, CSM-Net: Automatic joint segmentation of intima-media complex and lumen in carotid artery ultrasound images, Computers in Biology and Medicine, Volume 150, 2022, 106119, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2022.106119.

[5] Lucas Gago, Maria del Mar Vila, Maria Grau, Beatriz Remeseiro, Laura Igual, An end-to-end framework for intima media measurement and atherosclerotic plaque detection in the carotid artery, Computer Methods and Programs in Biomedicine, Volume 223, 2022, 106954, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2022.106954.

[6] Meiburger KM, Zahnd G, Faita F, Loizou CP, Carvalho C, Steinman DA, Gibello L, Bruno RM, Marzola F, Clarenbach R, Francesconi M, Nicolaides AN, Campilho A, Ghotbi R, Kyriacou E, Navab N, Griffin M, Panayiotou AG, Gherardini R, Varetto G, Bianchini E, Pattichis CS, Ghiadoni L, Rouco J, Molinari F. Carotid Ultrasound Boundary Study (CUBS): An Open Multicenter Analysis of Computerized Intima-Media Thickness Measurement Systems and Their Clinical Impact. Ultrasound Med Biol. 2021 Aug;47(8):2442-2455. doi: 10.1016/j.ultrasmedbio.2021.03.022. Epub 2021 Apr 30. PMID: 33941415.

[7] G. Zahnd, K. Kapellas, M. van Hattem, A. van Dijk, A. Sérusclat, P. Moulin, et al. A fully-automatic method to segment the carotid artery layers in ultrasound imaging: application to quantify the compression-decompression pattern of the intima-media complex during the cardiac cycle. Ultrasound Med. Biol., 43 (2017), pp. 239-257, 10.1016/J.ULTRASMEDBIO.2016.08.016

[8] Kristen M. Meiburger, Francesco Marzola, Guillaume Zahnd, Francesco Faita, Christos P. Loizou, Nolann Lainé, Catarina Carvalho, David A. Steinman, Lorenzo Gibello, Rosa Maria Bruno, Ricarda Clarenbach, Martina Francesconi, Andrew N. Nicolaides, Hervé Liebgott, Aurélio Campilho, Reza Ghotbi, Efthyvoulos Kyriacou, Nassir Navab, Maura Griffin, Andrie G. Panayiotou, Rachele Gherardini, Gianfranco Varetto, Elisabetta Bianchini, Constantinos S. Pattichis, Lorenzo Ghiadoni, José Rouco, Maciej Orkisz, Filippo Molinari,Carotid Ultrasound Boundary Study (CUBS): Technical considerations on an open multi-center analysis of computerized measurement systems for intima-media thickness measurement on common carotid artery longitudinal B-mode ultrasound scans, Computers in Biology and Medicine,Volume 144,2022,105333,ISSN 0010-4825,https://doi.org/10.1016/j.compbiomed.2022.105333.

[9] Molinari F, Meiburger KM, Saba L, Acharya UR, Ledda G, Zeng G, Ho SY, Ahuja AT, Ho SC, Nicolaides A, Suri JS. Ultrasound IMT measurement on a multi-ethnic and multi-institutional database: our review and experience using four fully automated and one semi-automated methods. Comput Methods Programs Biomed. 2012 Dec;108(3):946-60. doi: 10.1016/j.cmpb.2012.05.008. Epub 2012 Jun 1. PMID: 22658832.

[10] Meiburger, Kristen M, Filippo Molinari, U Rajendra Acharya, Luca Saba,Paulo Rodrigues, William Liboni, Andrew Nicolaides, and Jasjit S Suri, 2011, "Automated Carotid Artery Intima Layer Regional Segmentation, "Physics in Medicine & Biology 56 (13): 4073, https://doi.org/10.1088/0031-9155/56/13/021.

[11] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. arXiv. https://doi.org/10.48550/arXiv.1506.01497

[12] Jiang H, Learned-Miller E. Face detection with the faster R-CNN. 2017 12th IEEEinternational conference on automatic face & gesture recognition (FG 2017), IEEE; 2017, p. 650–7.

[13] MONAI Consortium. (2022). MONAI: Medical Open Network for AI (1.1.0). Zenodo. https://doi.org/10.5281/zenodo.7459814

[14] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., … Chintala, S. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Advances in Neural Information Processing Systems 32 (pp. 8024–8035). Curran Associates, Inc. Retrieved from http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[15] Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, Detectron2. https://github.com/facebookresearch/detectron2, 2019. [Online]. Available: https://github.com/facebookresearch/detectron2
[16] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin, "Albumentations: Fast and Flexible Image Augmentations," Information, vol. 11, no. 2, 2020, doi: 10.3390/info11020125.
[17] P. Iakubovskii, Segmentation Models Pytorch. GitHub, 2019. [Online]. Available: https://github.com/qubvel/segmentation_models.pytorch