# 2D and 3D Segmentation of uncertain local collagen fiber orientations in SHG microscopy

Lars Schmarje[1], Claudius Zelenka[1], Ulf Geisen[2], and Claus-C. Glüer[2] Reinhard Koch[1]

[1] Multimedia Information Processing Group, Kiel University, Germany
{las,cze,rk}@informatik.uni-kiel.de
[2] Molecular Imaging North Competence Center, Kiel University, Germany
{ulf.geisen,glueer}@rad.uni-kiel.de

**Abstract.** Collagen fiber orientations in bones, visible with Second Harmonic Generation (SHG) microscopy, represent the inner structure and its alteration due to influences like cancer. While analyses of these orientations are valuable for medical research, it is not feasible to analyze the needed large amounts of local orientations manually. Since we have uncertain borders for these local orientations only rough regions can be segmented instead of a pixel-wise segmentation. We analyze the effect of these uncertain borders on human performance by a user study. Furthermore, we compare a variety of 2D and 3D methods such as classical approaches like Fourier analysis with state-of-the-art deep neural networks for the classification of local fiber orientations. We present a general way to use pretrained 2D weights in 3D neural networks, such as Inception-ResNet-3D a 3D extension of Inception-ResNet-v2. In a 10 fold cross-validation our two stage segmentation based on Inception-ResNet-3D and transferred 2D ImageNet weights achieves a human comparable accuracy.

**Keywords:** comparison 2D and 3D · weight transfer from 2D to 3D · osteogenesis imperfecta · second harmonic generation · uncertain borders · rough semantic segmentation

## 1  Introduction

In a variety of medical issues and research activities, computed tomography (CT) scans are used for bone examinations. However, most CT scans only have a resolution in the millimeter range. Special CT procedures allow resolutions of a few µm [4,17]. For this reason single collagen fiber bundles of about 2-3 µm [1] can not be detected well in CT scans. The structure and orientation of these bundles allow us to make conclusions about changes in the bone (e.g. by growth or disease) [1]. These characteristics of the inner bone structure are valuable for research in the fields of age determination, disease detection and cancer research.

Second harmonic generation (SHG) microscopy can visualize these structures of collagen due to its higher resolution. This methods allows us to generate large

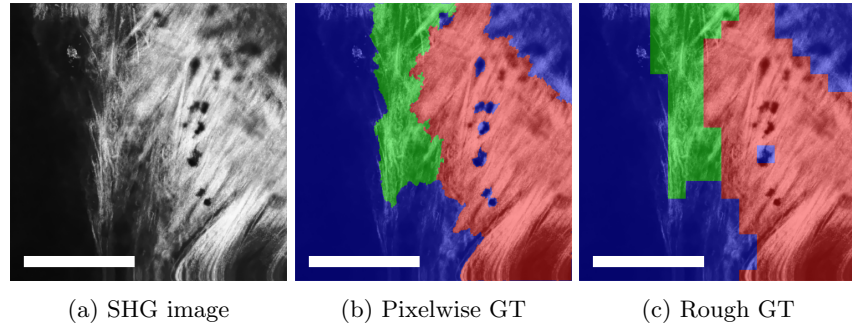(a) SHG image          (b) Pixelwise GT          (c) Rough GT

Fig. 1: Example of a SHG input image and the corresponding ground-truth - Note that while (b) is a pixel-wise ground-truth annotation we are in a setting with uncertain borders. Hence we should try to recreate a rough segmentation as shown in (c). The white scale bar depicts a size of 100 µm. Color code: Similar orientation - red, Dissimilar orientation - green, Not of interest - blue

amounts of dense 3D scans of collagen fibers in bones. It is time-consuming to create statistics of fiber bundles orientations or to mark regions of interest by experts. Moreover, manual annotations for large datasets are not feasible due to time constraints, budget and subjective biases. These biases are results of the uncertain borders in local fiber orientations. Therefore, an automatic analysis of fiber bundles orientation in large amounts of SHG data would benefit a variety of medical research activities. Large scale studies are not practical without automatic analysis.

The disease osteogenesis imperfecta (OI), also known as brittle bone disease, changes the orientation of fiber bundles in the bones of affected people and animals[1]. Hence SHG data of healthy and diseased mice is predestined for the evaluation of new methods.

Consequently, we developed different automatic algorithms for fiber bundle orientation analysis. We considered the orientation of fiber bundles in a local region as a classification problem. We focused ourselves on three classes: Bundles with similar (S) and dissimilar (D) orientations and everything else or not of interest (N) (e.g. noise, background). Figure 1 shows an example input and the corresponding ground-truth segmentation images.

The classification of a local region can also be defined as a rough semantic segmentation of the entire image. Since borders of local orientations are not well defined and can be described as highly uncertain and fluent, the goal is not to create pixel-wise segmentation. Therefore, we are interested in rough localization of these regions and their classification. Due to large regions which are not of interest (N) a large class imbalance in favor of this class exists and must be addressed. We analyzed the effect of uncertain borders on human performance in the task of rough fiber orientation segmentation by a user study.

We investigated classical approaches like Fourier analysis and state-of-the-art methods like deep neural networks. Instead of reporting only the best results we present a complete overview and comparison for future research in rough semantic segmentation. Most of our neural networks use a state-of-the-art backbone and domain specific adaptions. We aimed to change as little as possible in the original backbones to allow interchangeability with other backbones in the future. We show how a state-of-the-art 2D backbone can be used in 3D rough semantic segmentation. We call this network Inception-ResNet-3D. Especially we present a way to transfer pretrained 2D weights into the 3D case. The code is publicly available for reproducibility. [1]

To sum up, the main contributions of our work are:

- We report a systematical comparison of algorithms for classification and segmentation in 2D and 3D with uncertain borders. We use a novel dense 3D SHG dataset with more than 4500 slices for method development and testing. Human performance to classify uncertain collagen fiber orientations on this dataset is also reported.
- We show a general way to convert weights from the 2D into 3D.
- Our two stage approach with Inception-ResNet-3D and transferred 2D weights achieves a performance comparable to humans for uncertain collagen fiber segmentation in a 10 fold cross-validation.

## 2   Related Work

Currently neural networks are state-of-the-art in the field of image data classification (e.g. ImageNet [18]). A variety of neural networks have emerged over the years [10,21,7,8,24,23]. These networks started with a simple architecture (e.g. VGG-16 [21]). They integrated new structure elements like residual [7] and inception blocks [24] as they were developed and proved their superior performance. This development led to an increase of the top1 accuracy on the ImageNet test set from 71.3% with VGG-16 to 80.3% with Inception-ResNet-v2. Parallel the depth and thereby the complexity increased from 23 to 572 layers [2].

Semantic segmentation gives a classification for every pixel in an image and is an extension of a classification problem. Shelhamer et al. [20] first proposed to use fully convolutional networks to solve semantic segmentation. U-Net [16] is a network for semantic segmentation which was designed for medical images. Often semantic segmentation networks consist of a down- and a upsampling part [20,16].

However, the current state-of-the-art approaches for image classification and semantic segmentation have two major drawbacks in the context of uncertain local fiber orientation classification. We have 3D data and a high uncertainty for the borders. Most research focuses on 2D data while Zhou et al. [26] showed

---

[1] https://github.com/Emprime/uncertain-fiber-segmentation
[2] Values are based on the reference implementation in Keras

that it is beneficial to use the 3D information for organ segmentation. Networks like PointNet [14] can classify 3D point clouds yet they do not consider dense 3D input as we have. The network 3D-U-Net [3] represents an expansion of U-Net to 3D data. It is typically used to segment 3D objects like organs [3]. This fixes the first drawback while the second one remains. Objects with uncertain borders like our fiber orientations are not well represented.

While 3D extensions of Inception-ResNet-v2 have been presented in [9,6] the usage of 2D pretraining is not so widely used. Parallel to our research Shan et al. proposed a 2D weight transfer strategy to 3D [19] which is most similar to ours (see subsection 4.1).

Collagen structures in SHG images have been analyzed in several publications [2,25,5,1,22,15,11]. They were analyzed in tissue [2] and bones [5,25]. Rao et al. [15] presented how Fourier analysis can be used to investigate the orientation of collagen fibers. The Fourier analysis was extended from small regions to the whole scan in [1,22,11]. The analysis classified small image parts as anisotropic, isotropic and dark. These classifications where used to calculate the distributions of classes over an image. In [22] these distributions where used to detect injured tendons. Moreover, Ambekar et al. [1] showed the change of distribution due to aging can be used to determine the age of pigs. Lau et al. [11] used the 3D information of SHG data and could show an increase in performance.

Nevertheless, their analysis is based on only few (<100) images. An analysis on larger amounts of data is not known. The data shown in the papers seems to be of overall of a good quality. Artifacts, noise and blurring and impact of performance was not reported.

Liang et al. [12] state to be the first to analyze SHG images with neural networks. They estimated the elastic properties of collagenous tissue. A classification or segmentation of fibers were not part of their investigation.

To our knowledge, we are the first who use neural networks to automatically classify and segment local collagen fiber orientation in large amounts of 2D or 3D data. In contrast to previous neural network literature we use 3D data and adapt our networks to uncertain borders. In comparison to earlier fiber analyses we utilize neural networks to process large amount of mixed quality data.

## 3   User study

While we knew that we operated in a context with uncertain borders we did not know how this would impact performance. Therefore, we investigated this issue by a random sample user study. Our goal was to examine how well humans can classify and segment local fiber orientations. We compared 15 different people with each other (interpersonal) and 5 results of the same person over time (intrapersonal).

The participants were given two tasks. The first task was to chose one annotation out of 5 given example annotations for one image. This task was repeated for 10 different images. The second task was to create an annotation for 24 images.

For the first task we calculated the Pearson correlation coefficient between the annotation selections of all participants (interpersonal) or over time (intrapersonal). This leads to a mean absolute coefficient of 0.44 with a standard deviation of 0.26 for the intrapersonal comparison. The interpersonal comparison results in a mean absolute coefficient of 0.24 with a standard deviation of 0.2.

For the second task we calculated the accuracies of the created annotations with the ground-truth (see subsection 5.2 for the metric definition). The intrapersonal comparison reached a mean accuracy of 78.29% with a standard deviation of 2.40% over all 24 images. The interpersonal comparison resulted in a mean accuracy of 58.83% with a standard deviation of 7.44%.

All in all we see that it is more difficult for different people to select or create consistent annotations than for one person over time. However, even for a person over time the selection and creation is not perfect. We can state that humans achieve only about 78% accuracy consistency with themselves. If we train and evaluate a neural network on human created ground-truth with this consistency rate we can not expect that an algorithm performs significantly better.

## 4   Methods

All methods use the same datasets although the 2D methods ignore the inherent three-dimensional information. Therefore, 2D data will be referred to as scan slice or image and 3D data as scan. For all methods we investigated a variety of hyperparameters such as batchsize, backbones and loss variations. We will mention in the method description only important hyperparameter selections and specialties. For further details see the supplementary materials.

### 4.1   Weight transfer 2D to 3D

We want to utilize the pretrained ImageNet weights in our 3D Networks and thus we have to transfer the 2D kernel weights into 3D kernel weights. Technically this is a function $I : \mathbb{R}^{w \times h \times c} \to \mathbb{R}^{w \times h \times d \times c}$ that transforms a 3D (width, height, channels) matrix $M_1$ into a 4D (width, height, depth, channels) matrix $M_2$ with $w, h, d, c \in \mathbb{N}$ .

We investigated two methods for the weight transformation. We denote the set $\{1, .., N\}$ by $[n]$ . The first approach is to divide $M_1$ by $d$ and stack them $d$ times to create $M_2$ for a given depth $d \in \mathbb{N}$:

$$M_{2(i,j,k,l)} = M_{1(i,j,l)}/d \text{ for all } i \in [w], j \in [h], k \in [d], \text{ and } l \in [c]. \qquad (1)$$

The second approach is to insert the 3D matrix into the 4D matrix and fill the

rest up with zeros. Shan et al. proposed in [19] a similar method. For given odd depth $d \in \mathbb{N}$ and center element $\hat{c} = \frac{(d-1)}{2} + 1$ this is defined as

$$M_{2(i,j,k,l)} = \begin{cases} M_{1(i,j,l)} & \text{for } k = \hat{c} \\ 0 & \text{for } k \in [d] \setminus \{\hat{c}\} \end{cases} \tag{2}$$

and all $i \in [w], j \in [h], k \in [d], m \in [c]$.

Henceforth, we refer to theses transformations if we talk about 2D weights in a 3D context or transferred weights. See subsection 4.4 for further details on the selected transfer strategies. We use this weight transformations in the network Inception-Resnet-3D a 3D extension of Inception-ResNet-v2 [23]. In general the architecture is the same but with 3D layers as done before by [9,6]. In the case of asymmetric input data we introduce asymmetric strides in the downsamplings in the stem block. These asymmetric strides create symmetric input for deeper layers.

### 4.2   Weighted Focal Loss

As mentioned before we have to address the issue of class imbalance in our training data and chose to investigate different loss variations. Lin et al. defined the novel loss function Focal Loss in [13] which should automatically balance the contribution of classes in skewed cases. As mentioned in [13] the loss $L : [0,1]^n \times \{0,1\}^n \to [0,1]$ can be extended to the non binary case as shown in Equation 3 with $n \in \mathbb{N}$ being the number of classes and $\gamma$ the Focal Loss parameter. Furthermore, we integrated weights $w \in \mathbb{R}^n$ for every prediction $\hat{y} \in [0,1]^n$ and ground-truth $y \in \{0,1\}^n$. In our case $n = 3$ and the values of $y$ and $\hat{y}$ correspond to these three classes.

$$L(\hat{y}, y) = -\sum_{i \in [n]} w_i (1 - \hat{y}_i)^\gamma y_i log(\hat{y}_i) \tag{3}$$

Keep in mind that Equation 3 is equal to cross entropy if $\gamma = 0$ and $w = \{1\}^3$.

### 4.3   2D methods

**Fourier analysis** As a reference we reimplemented a classification based on the Fourier analysis in small image patches [15]. We used thresholds like in [15] to discriminate different classes.

**Classification** A straight forward approach for local region classification is to create such regions by splitting the images into smaller image parts each with the same width and height. An equal class distribution was enforced on these image regions and a graphical representation of the input is given in Figure 2a. A rough segmentation for the image can be generated out of the classifications for all image parts.

(a) 2D classification

(b) 2D semantic segmentation

(c) 3D combination of 2D classifications

(d) 3D classification

(e) 3D two stage segmentation - first stage
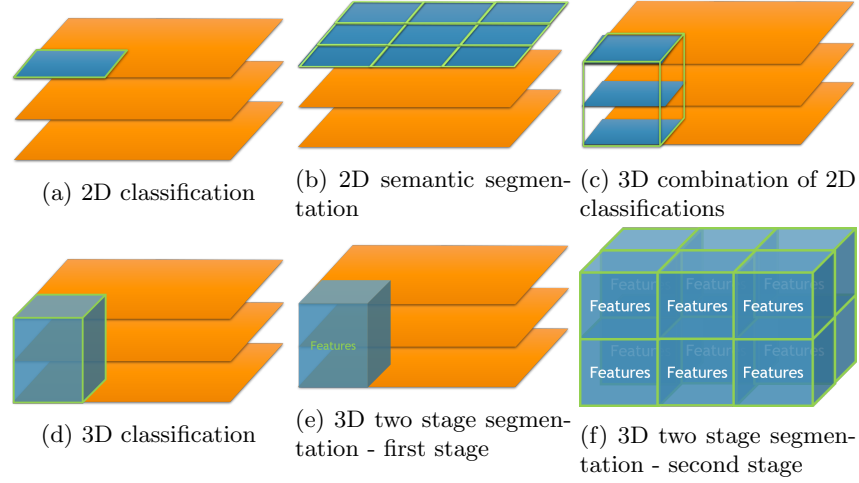
(f) 3D two stage segmentation - second stage

Fig. 2: Graphical representation of the used input and output for the different proposed methods - The orange slices represent the SHG images which form a scan together. The blue tiles or blocks are the inputs to the network while the green markings show the output format. For example (a) shows a 2D image part as an input and one value as output. While (f) takes features as a 3D matrix and outputs a 3D matrix.

During development we discovered that classification accuracy was higher if the image part size was larger. These larger sizes lead to a rougher segmentation and thus we used an ensemble and majority voting to combine the benefits of a smaller image part size and the higher accuracy of larger image parts. We used an Inception-ResNet-v2 [23] backbone with pretrained weights, $\gamma = 0$ and $w = \{1\}^3$.

**Semantic segmentation** Classification of image parts has two major drawbacks. It is time consuming since a lot of image parts have to be processed and a post processing step is needed to combine the classifications to a segmentation. A parallel rough semantic segmentation of an image can overcome both these problems. In contrast to other literature [20,16] we use only a downscaling part and not an upsampling part in our segmentation network. As described earlier we are not interested in a fine segmentation and can drop the upsampling because of this (see Figure 2b).

The segmentation networks differ from the original backbones mostly in the output. The architecture is shown in Figure 3a. We use an average pooling layer and a 1x1 convolutional layer instead of a global average pooling layer and a fully connected layer for multiple reasons. Firstly, we want to create a matrix as an output which consists of softmax outputs for every row and column. Secondly, we can incorporate the neighborhood information through the pooling layer.

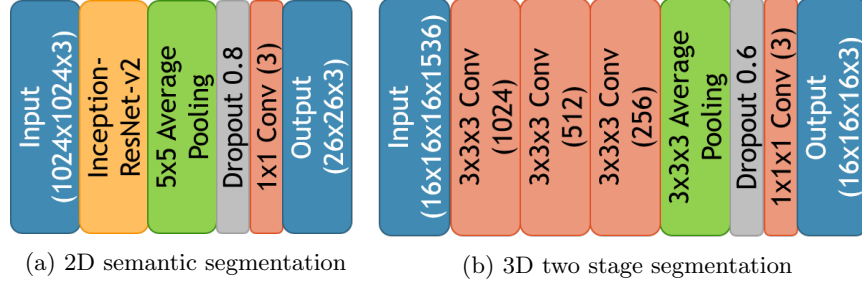(a) 2D semantic segmentation    (b) 3D two stage segmentation

Fig. 3: Architecture of the 2D semantic segmentation network with backbone and the second stage network for the 3D two stage segmentation - Different layers have different colorings. The main part of Inception-ResNet-v2 up to the global average pooling layer is described as one block. The dimensions for in- and output and the number of feature maps for convolutional layers are given in brackets. The kernel size is given before the layer name.

Thirdly, we have to use a convolutional layer for the output because otherwise the number of parameters for the fully connected layer would have become unmanageable. In addition we can create a finer segmentation than the ensemble above (subsection 4.3) and also use neighborhood information due to average pooling layers. We used an Inception-ResNet-v2 [23] backbone with pretrained weights, $\gamma = 0$ and $w = (\frac{16/}{41}, \frac{24}{41}, \frac{1}{41})$. The weights are needed to account for the class imbalance in the input data.

### 4.4  3D methods

**Combination of 2D classifications** This method is an extension of the 2D classification by combination. We use the 2D classifications and include 3D information by averaging over the classifications which are positioned next to each other. This simple aggregation yields a 3D classification but inherits the drawbacks of the 2D classification. A graphical representation is shown in Figure 2c.

**Classification** This method is a extension of the 2D classification in 3D. Instead of image parts (square) we use scan blocks (cuboid) for the classification. The graphical representation of the input and output is shown in Figure 2d. However, we do not use an ensemble to combine different 3D classifications. We used our proposed Inception-ResNet-3D with transferred 2D weights based on ImageNet, $\gamma = 2$, $w = \{1\}^3$ and used the transfer strategy based on Equation 1.

**Two stage segmentation** In the 2D case parallel segmentation of a complete image could utilize the neighborhood information for every entry in the output matrix. In order to combine the information of a whole scan for an output and still fit in the memory of one GPU we had to take a two stage approach. The idea

is to extract features with a pretrained network and then combine these features in a second network to create a 3D matrix where every entry correspondence to the three classes (graphical representation see Figure 2e and Figure 2f).

We used Inception-ResNet-3D as an extraction network with transferred ImageNet weights with the transfer strategy based on Equation 1. Unlike in 3D classification we do not want one classification but the features as output. The second stage is a small network out of convolutional and average pooling layer to combine the 3D matrix of features to class predictions. The architecture is shown in Figure 3b and was inspired by Figure 3a.

## 5 Experimental Results

### 5.1 Dataset

We developed our methods on one dataset which was created by the MOIN CC[1]. The dataset consists of 4736 SHG images from 35 scans of 6 mice where 3 mice had the disease OI and the others do not. The scans were taken on different parts of the legs and had a resolution of 1000 x 1000 px or 1024 x 1024 px while capturing 250 µm x 250 µm of the bone. We cannot downscale these images because we would loose the necessary resolution fine fiber structures. The depth of each scan was variable and ranged from 78 to 214 images while the distance in the bone between each image is 0.5 µm.

A main property of the data is the class imbalance. Roughly 2% of the data belongs to the class S (similar orientation) and 3% to the class D (dissimilar orientation). The remaining 95% belong to the class N (not of interest) and are, therefore, not interesting in medical research. The data was split in to a training, validation and test set. Figure 4 displays three example of used SHG images which represent the variety in the input data.

Moreover, investigations of selected background regions showed a high scanner noise. The average grey value (0-255) of the background should be zero but varied between 2.91 and 40.2. On average the registered grey values differ from the real values with a mean of 9.18 and a standard deviation of 8.79.

### 5.2 Evaluation Metrics

We use an adapted accuracy function to measure the performance of our results. In short we use the mean of accuracies per class as our accuracy measure. Our function $meanacc$ for prediction $\hat{y} \in [n]^k$ and ground-truth $y \in [n]^k$ with $n \in \mathbb{N}$ number of classes and $k \in \mathbb{N}$ entries is defined as follows

$$meanacc(\hat{y}, y) = \frac{1}{n} \sum_{j=1}^{n} \frac{\sum_{i=1}^{k} \mathbb{1}_{\hat{y_i}=y_i, \hat{y_i}=j}}{\sum_{i=1}^{k} \mathbb{1}_{y_i=j}}. \tag{4}$$

---

[1] **Mo**lecular **I**maging **N**orth **C**ompetence **C**enter

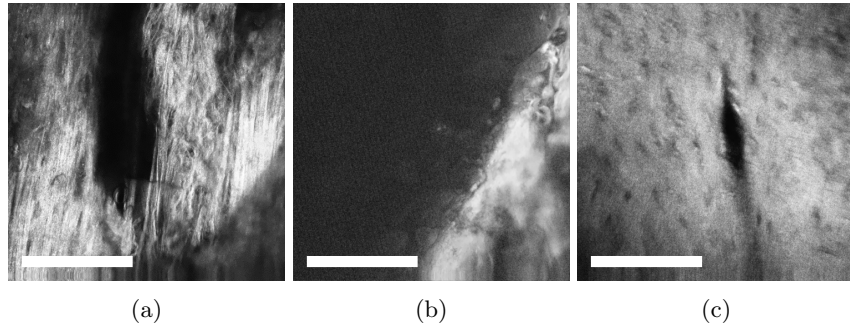(a)                          (b)                          (c)

Fig. 4: Examples of used SHG images - (a) shows a desirable image. We have sharp and fine collagen structures while the noise ratio is low. (b) shows a noisy image where we can see collagen structures on the right hand side. Due to blurring and noise we can only detect rough shapes. (c) shows the macroscopic structure of the bone. We can detect the hole in the center but no single fibers or fiber bundles. The white scale bar depicts a size of 100 μm.

The function *meanacc* has the benefit of being stable against class imbalance and is the same as the normal accuracy in the case of class balance. It allows an estimation of performance in a single value without tuning weights. In this paper accuracy on our data always refers to *meanacc*.

## 5.3   Method comparison

We used a strict data separation during development to be able to compare methods. All hyperparameters were selected based on the validation set. The test set was only used during method comparison. In general we noticed two trends. Firstly, better network performance on ImageNet translates to improved accuracies in all our methods. This isn't noteworthy for tasks like 2D classification but for improvements in segmentation and feature extraction tasks it is. Secondly, pretrained weights ensure a good initialization and lead to greater accuracies. This is expected and reported for 2D classification tasks but the fact that pretraining can be interpolated to a 3D case and still ensures greater performance is significant.

Table 1 compares all presented methods with regard to run-time, resolution and accuracy. The Fourier analysis results in the worst performance even on the validation set. Due to this inferior performance and the long run-time we did not evaluate the method further. We believe that the high variability in the data can not be captured from such a simple approach. The method of 2D semantic segmentation has the fastest run-time and the finest resolution. The accuracy is with about 65% the best for all 2D methods. The methods 3D classification and 3D two stage segmentation score a higher accuracy but have a rougher resolution and a longer run-time. The best accuracy of 72% is achieved by 3D two stage segmentation.

| Method | Run-time | Resolution | Accuracy |
|---|---|---|---|
| 2D Fourier analysis | N/A* | 16 x 16 x 1 | 55.00%* |
| 2D classification | 101 min | 64 x 64 x 1 | 59.54% |
| 3D combination of 2D classifications | 101 min | 64 x 64 x 16 | 59.68% |
| 2D semantic segmentation | **14 min** | **40 x 40 x 1** | 64.58% |
| 3D classification | 17 min | 128 x 128 x 64 | 70.51% |
| 3D two stage segmentation | 72 min | 64 x 64 x 16 | **72.18%** |

Table 1: Overview of the best results on the test set for each method - Run-time is the time it took to process the test set once which includes pre- and postprocessing. Resolution describes the number of pixels in the input that are mapped to one output value. Smaller resolutions result in finer segmentation but also in an accuracy drop for some methods and are, therefore, not reported here. The accuracy is reported on the test set. The best result of each column is marked bold. *Due to the inferior performance on the validation data and long run-time we did not evaluate the method Fourier analysis on the test set.

In general we see that it is beneficial to process as much data as possible simultaneously to achieve a high accuracy. This result can be explained due to the fact that simultaneous processing incorporates a larger neighborhood. Furthermore, we see that there is not one best method in all regards. Only a trade-off between different characteristics can be chosen.

It is remarkable that the features used in the second stage were extracted with transferred 2D ImageNet weights and still lead to the best results. Furthermore no adaption to domain specific weights is needed. In the context of a human consistency of about 78.29% in the user study (section 3) a result of 72% is remarkable.

### 5.4   Cross-validation

We did 10 fold cross-validation on the complete dataset to verify that the chosen random split in different sets introduced no bias and represent the real data distribution. The data was split 10 times into a training (50%), a validation (25%) and a test (25%) set. We split randomly but kept only splits where at least 2% of the data had the class S and another 2% the class D. We put scans from the same bone region into the same set.

We trained the method two stage segmentation on the training set, used the best weights on the validation set and evaluated on the test set. The mean accuracy over 10 runs is 75.79%. Figure 5 shows that the accuracies for all runs are in a margin of about 10% around the mean. Some runs achieve an accuracy above the expected accuracy based on the user study.
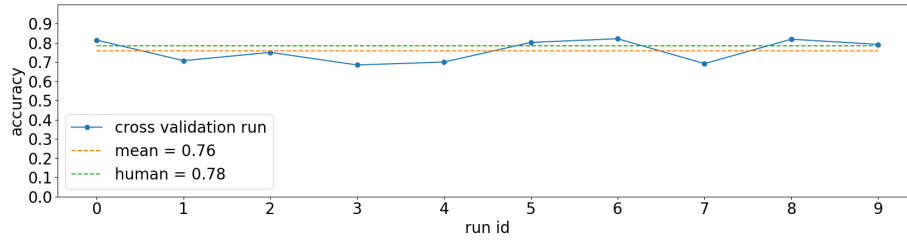
Fig. 5: Results of the cross-validation with mean and human performance based on the user study

## 6   Conclusion

We compared a variety of methods for rough semantic segmentation of collagen fiber orientation in 2D and 3D. As a dataset we used a novel collection of dense 3D SHG scans which is larger and more diverse as previously used datasets [1,11]. Our conducted user study implies that human can reach an average consistency of about 78.29% on the task of uncertain collagen fiber orientation segmentation. This results in a similar expected accuracy for trained algorithm due to the human annotated ground-truth. We showed how to use transformed 2D ImageNet weights in 3D networks in general and in Inception-ResNet-3D in particular. We proposed a two stage model that can simultaneously process large 3D inputs and use transformed 2D weights. This best method two stage segmentation achieves an average accuracy of 75.79% over 10 fold cross-validation. Based on the user study we can say that we created an algorithm with near human performance.

The presented user study led to great insights into possible performance of neural networks. It will be beneficial to repeat the user study at a larger scale. We are confident that two stage segmentation with transferred weights can be applied in different 3D classification and rough segmentation tasks. We will investigate these usages in the future. Furthermore, we will investigate how to create more objective ground-truth for example by leveraging pretrained features and reduced supervision.

# References

1. Ambekar, R., Chittenden, M., Jasiuk, I., Toussaint, K.C.: Quantitative second-harmonic generation microscopy for imaging porcine cortical bone: comparison to sem and its potential to investigate age-related changes. Bone **50 3**, 643–50 (2012)
2. Campagnola, P.J., Loew, L.M.: Second-harmonic imaging microscopy for visualizing biomolecular arrays in cells, tissues and organisms. Nature Biotechnology **21**, 1356–1360 (2003)
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)
4. Genant, H., Engelke, K., Prevrhal, S.: Advanced ct bone imaging in osteoporosis. Rheumatology **47**(suppl_4), iv9–iv16 (2008)
5. Genthial, R., Beaurepaire, E., Schanne-Klein, M.C., Peyrin, F., Farlay, D., Olivier, C., Bala, Y., Boivin, G., Vial, J.C., Débarre, D., et al.: Label-free imaging of bone multiscale porosity and interfaces using third-harmonic generation microscopy. Scientific reports **7**(1), 3419 (2017)
6. Hassani, B., Mahoor, M.H.: Facial expression recognition using enhanced deep 3d convolutional neural networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW) pp. 2278–2288 (2017)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 770–778 (2016)
8. Huang, G., Liu, Z., van der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 2261–2269 (2017)
9. Kang, G., Liu, K., Hou, B., Zhang, N.: 3d multi-view convolutional neural networks for lung nodule classification. PloS one **12**(11), e0188290 (2017)
10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Commun. ACM **60**, 84–90 (2012)
11. Lau, T.Y., Ambekar, R., Toussaint, K.C.: Quantification of collagen fiber organization using three-dimensional fourier transform-second-harmonic generation imaging. Optics express **20 19**, 21821–32 (2012)
12. Liang, L., Liu, M., Sun, W.: A deep learning approach to estimate chemically-treated collagenous tissue nonlinear anisotropic stress-strain responses from microscopy images. Acta biomaterialia **63**, 227–235 (2017)
13. Lin, T.Y., Goyal, P., Girshick, R.B., He, K., Dollár, P.: Focal loss for dense object detection. 2017 IEEE International Conference on Computer Vision (ICCV) pp. 2999–3007 (2017)
14. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 77–85 (2017)
15. Rao, R.A.R., Mehta, M.R., Toussaint, K.C.: Fourier transform-second-harmonic generation imaging of biological tissues. Optics express **17 17**, 14534–42 (2009)
16. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
17. Rueckel, J., Stockmar, M.K., Pfeiffer, F., Herzen, J.: Spatial resolution characterization of a x-ray microct system. Applied radiation and isotopes : including data,

instrumentation and methods for use in agriculture, industry and medicine **94**, 230–234 (2014)

18. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M.S., Berg, A.C., Fei-Fei, L.: Imagenet large scale visual recognition challenge. International Journal of Computer Vision **115**, 211–252 (2015)

19. Shan, H., Zhang, Y., Yang, Q., Kruger, U., Kalra, M.K., Sun, L., Cong, W., Wang, G.: 3-d convolutional encoder-decoder network for low-dose ct via transfer learning from a 2-d trained network. IEEE Transactions on Medical Imaging **37**, 1522–1534 (2018)

20. Shelhamer, E., Long, J., Darrell, T.: Fully convolutional networks for semantic segmentation. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 3431–3440 (2015)

21. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR **abs/1409.1556** (2015)

22. Sivaguru, M., Durgam, S.S., Ambekar, R., Luedtke, D., Fried, G.A., Stewart, A.W., Toussaint, K.C.: Quantitative analysis of collagen fiber organization in injured tendons using fourier transform-second harmonic generation imaging. Optics express **18 24**, 24983–93 (2010)

23. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)

24. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S.E., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp. 1–9 (2015)

25. Thomas, B., McIntosh, D., Fildes, T., Smith, L., Hargrave, F., Islam, M., Thompson, T., Layfield, R., Scott, D., Shaw, B., et al.: Second-harmonic generation imaging of collagen in ancient bone. Bone reports **7**, 137–144 (2017)

26. Zhou, X., Yamada, K., Kojima, T., Takayama, R., Wang, S., Zhou, X., Hara, T., Fujita, H.: Performance evaluation of 2d and 3d deep learning approaches for automatic segmentation of multiple organs on ct images. In: Medical Imaging 2018: Computer-Aided Diagnosis. vol. 10575, p. 105752C. International Society for Optics and Photonics (2018)