Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

ACM BCB. Author manuscript; available in PMC 2021 May 26.

Published in final edited form as:

ACM BCB. 2019 September ; 2019: 299–306. doi:10.1145/3307339.3342153.

SAU-Net: A Universal Deep Network for Cell Counting

Yue Guo,
University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Guorong Wu,
University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Jason Stein,
University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Ashok Krishnamurthy
Renaissance Computing Institute, Chapel Hill, NC, USA

Abstract

Image-based cell counting is a fundamental yet challenging task with wide applications in biological research. In this paper, we propose a novel Deep Network designed to universally solve this problem for various cell types. Specifically, we first extend the segmentation network, U-Net with a Self-Attention module, named SAU-Net, for cell counting. Second, we design an online version of Batch Normalization to mitigate the generalization gap caused by data augmentation in small datasets. We evaluate the proposed method on four public cell counting benchmarks - synthetic fluorescence microscopy (VGG) dataset, Modified Bone Marrow (MBM) dataset, human subcutaneous adipose tissue (ADI) dataset, and Dublin Cell Counting (DCC) dataset. Our method surpasses the current state-of-the-art performance in the three real datasets (MBM, ADI and DCC) and achieves competitive results in the synthetic dataset (VGG). The source code is available at https://github.com/mzlr/sau-net.

Keywords

cell counting; neural networks; data augmentation

1 INTRODUCTION

Image-based cell counting is essential to a wide-range of biological research problems [4, 22], since variability in the number of cells of a given type between wild-type and disorder-associated mutant animal models can provide valuable insights into the underlying cellular mechanism of those disorders. For example, the count of proliferative neural progenitor cells

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

yueguo@cs.unc.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

can provide a quantitative metric of the cellular basis underlying observed macrocephaly in individuals with mutations in Chromodomain Helicase DNA binding protein 8 (CHD8), which is also highly associated with Autism Spectrum Disorder [4]. Therefore, it is of great importance to provide a universal framework to perform cell-counting for various cell types imaged with various labels and microscopes.

Current public benchmarks for image-based cell counting [10, 13, 19, 21] use a single pixel to represent each cells, i.e., the center of a cell is set to 1 otherwise 0. This simple dot-annotation method is advantageous for this task since there could be hundreds of cells in a single images and a complete annotation would be time-consuming. Therefore, recent work approaches this task from a regression perspective, i.e., learning a mapping from an input image to either a scalar (cell count) [19, 29] or a density map [13, 18, 21, 28], from which the cell count can be inferred by integration. Inspired by the recent success of U-Net based methods for image segmentation [7, 20, 24] and the similarity between a density map and a segmentation map, we follow the latter direction and choose U-Net as the regression model.

One of the main obstacles in the learning of the mapping process is that the density maps are dot-annotated and thus extremely sparse, which makes the model difficult to train, as shown in Figure 1. Previous methods [13, 28] remedy this problem by applying a Gaussian kernel at each dot annotation to blur the density map. Built upon this work, we propose to use Self-Attention module [25] to learn long range, non-local dependencies in images, forcing the model to “focus” on the foreground pixels instead of background pixels.

Another challenge is the generalization gap when training with small datasets using Batch Normalization [9] in conjunction with data augmentation. Data augmentation methods such as random cropping or flipping, are crucial to improve model robustness and prevent over-fitting when limited training data is available [24, 26]. Recently proposed Batch Normalization has become an essential part in training deep networks to combat covariate shift [8]. By normalizing activations with each batch during training, it can help expedite the convergence of the optimization but it will also increase the generalization gap due to a biased estimation with limited data [26]. The situation deteriorates especially when used with data augmentation since data augmentation will introduce additional disturbance to the distributions. To address this challenge, we introduce an online version of Batch Normalization, which will re-normalize the activations using training and test data, instead of the moving mean and variance from training, during inference. This extension has a negligible impact on computation when implemented in parallel and overcomes the generalization gap without complicated heuristics.

To validate the effectiveness of the proposed method, we test on four public benchmarks for cell counting. State-of-the-art performance is obtained in the Modified Bone Marrow (MBM) dataset [10], human subcutaneous adipose tissue (ADI) dataset [21] and Dublin Cell Counting (DCC) datasets [19]. On the synthetic fluorescence microscopy (VGG) dataset [13], our result is on par with the leading method.

In summary:

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

- We propose a novel deep architecture, SAU-Net, to incorporate U-Net [24] with Self-Attention module [25] for cell counting. An online-version of Batch Normalization is also developed to solve the generalization gap for small datasets.
- SAU-Net outperforms the state-of-the-art methods in three out of four public cell counting benchmarks and achieves competitive results in the other synthetic benchmark, highlighting the universal nature of the proposed method.
- The source code is available to the research community at https://github.com/mzlr/sau-net, aimed at stimulating further investigations on this topic.

The rest of the paper is organized as follows: We will first review the related work for cell counting in Section 2 and then present the proposed method in Section 3. Section 4 will briefly describe the public benchmarks used in this study and Section 5 will compare our results with the state-of-the-art methods for each benchmark. Finally, Section 6 will conclude the paper with a discussion and future work.

2 RELATED WORK

There are generally two categories of cell counting methods: detection based [1, 3] and regression based [13, 18, 19, 21, 28]. The former approaches use a detector to localize each individual cell and the cell count can be obtained by counting the detected cells. Cell detection, however, remains a challenging task due to occlusion, shape variation, etc., and training such a model requires annotations for individual cells, which is time-consuming given the high cell density in images. Therefore, this paper focuses on the regression-based method.

Earlier regression based methods learned a mapping from dense local image features to a density map. Lempitsky and Zisserman [13] first used a linear regression with dense SIFT features to predict the density map. Later, a regression forest was proposed in [5] to replace the linear regression for better density map estimation. Arteta et al. [2] extended the pipeline by a local feature vocabulary and ridge regression in an interactive fashion.

Recent methods favor Deep Neural networks due to its versatility in various research areas, e.g., computer vision[11, 23], medical imaging [7, 30], natural language processing [25]. Xie et al. [28] relied on fully convolutional regression networks [14] to estimate the density map filtered by a Gaussian kernel. Cohen et al. [21] followed this direction but constructed the density map based on the receptive fields of the networks. They first filtered the dot annotations with a square kernel and then summed the value in a sliding window fashion, with each window corresponding to a receptive field in the network. This redundant counting improved the counting accuracy but could over-fit to the background in some cases. This is because the network no longer obtained the cell count by identifying individual cells in a density map and it could simply average the regression prediction when large areas of background exist in images.

Alternatively, Xue et al. [29] divided the input into multiple sub-images and tested multiple neural networks to map sub-images to a single scalar, namely the cell count. Similarly, [19]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

used a pre-train network to learn the same mapping for cross-domain counting. One drawback for patch-based counting is the miscalculation of cells on patch boundaries since the annotation is sparse while the actual cells can be very crowded.

To address the limitations mentioned above, we propose SAU-Net, a U-Net [24] with a Self-Attention module [25], which avoids the background over-fitting and boundary miscalculation problems. We also extend the Batch Normalization [9] to further improve the counting accuracy. Note that there is a similarity between our work and an attention based U-Net for image segmenting in [20], but our model differs in the way we incorporate the attention module; also our application is different.

3 METHOD

In this section we will first introduce the overview of our regression based cell counting method and then present our novel contributions.

3.1 Overview

The goal of our regression based cell counting method is to learn a mapping F from a $m \times n$ image I to a $m \times n$ density map D, denoted as

$$F: I \rightarrow D \left(I, D \in R^{m \times n}\right) \tag{1}$$

The density map D is obtained by placing a Gaussian kernel at each dot annotation representing individual cells. For simplicity, we choose a fixed bandwidth for all the Gaussian kernels. Finally the cell count can be computed by integration over the density map D.

3.2 SAU-Net

We propose a novel model, called SAU-Net, to represent the mapping F in Equation 1. SAU-Net incorporates the Self-Attention module [25] on top of a U-Net architecture [24], whose encoding-decoding structure proves suitable for various medical imaging tasks [7, 20, 24]. In addition, we add a Batch Normalization [9] layer after every convolution and deconvolution layer in U-Net. The overall structure of SAU-Net is illustrated in Figure 2.

Self-Attention can be viewed as a non-local weighted averaging operation in deep neural networks [27]. Mathematically, we have

$$Self-Attantion(x) = softmax\left(f(x)^T g(x')\right)h(x') \tag{2}$$

where $x \in R^{H \times W \times D}$ are the activations from the previous layers and used as inputs to the Self-Attention module. H, W and D represents the height, width and channel number of the activations, respectively. $x' \in R^{\frac{H}{2} \times \frac{W}{2} \times D}$ is obtained by sub-sampling x with 2×2 max-pooling for computation efficiency, following [27]. f, g and h are linear embeddings, implemented as 1×1 convolution and proper reshaping is performed for matrix

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

multiplication, i.e., $reshape : R^{H\times W\times D} \rightarrow R^{(HW)\times D}$. $softmax$ here works as a scaling function so that the weights for the averaging sum to 1, and $\otimes$ denotes matrix multiplication.

Self-Attention computes weights based on feature relationships, i.e., $softmax(f(x)^T g(x'))$ in Equation 2, across the whole image region, whereas conventional convolutional layers can only process local information, e.g., 3 × 3 convolution. By combining the Self-Attention module with convolutional layers in U-Net, SAU-Net can enjoy a richer hierarchy which can learn the mapping based on both local and global relationship. Given the encoding and decoding structure of U-Net, it is generally believed the deeper convolutional layer of U-Net has the more abundant feature information. Therefore, we choose to incorporate the Self-Attention module at end of the encoding path of U-Net for more accurate feature relationships.

3.3 Online Batch Normalization

Batch Normalization has become an integral part of modern deep networks [7, 8, 27]. It normalizes internal activations to help stabilize the distributions with mini-batches, enabling higher learning rates, faster convergence and training of deeper networks. We briefly describe Batch Normalization in Algorithm 1 to make this paper self-contained.

With limited training data, Batch Normalization could lead to over-fitting due to a biased estimation of μ and σ in Algorithm 1 during inference. Data augmentation would further increase increase the generalization gap since it adds extra variation into the data. In order to make this two techniques compatible with each other, we extend Batch Normalization with a simple modification, an online version. In particular, online Batch Normalization will re-normalize the activations using training and test data, instead of the moving mean and variance from training, during inference. Algorithm 2 presents our online Batch Normalization.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Algorithm 1: A single layer of Batch Normalization [9]

Input: A mini-batch B with size M of activations x: $B = \{x_1, \ldots, x_M\}$; trainable parameters: γ, β; constant for numerical stability: ϵ; current moving mean and standard deviation: μ, σ; moving average update rate: α; test samples with size N: $T_{test} = \{x_1, \ldots, x_N\}$

Training:

/* mini-batch mean and variance */

$\mu_B \leftarrow mean(\{x_1, \ldots, x_M\})$;

$\sigma_B \leftarrow std(\{x_1, \ldots, x_M\}) + \epsilon$;

/* normalize activations within the mini-batch */

for $i = 1 \ldots M$ **do**

  $y_i \leftarrow \gamma \frac{x_i - \mu_B}{\sigma_B} + \beta$

end

/* update moving averages */

$\mu \leftarrow \mu + \alpha(\mu_B - \mu)$;

$\sigma \leftarrow \sigma + \alpha(\sigma_B - \sigma)$;

optimize γ, β via back-propagation ;

Inference:

for $j = 1 \ldots N$ **do**

  $y_j \leftarrow \gamma \frac{x_j - \mu}{\sigma} + \beta$

end

Our extension is especially suitable when working with small datasets. For example, most datasets [10, 13, 19, 21] for cell counting have less than 100 training images. Since no back-propagation is needed for inference, these can be easily fit into the GPU memory for parallel computation, thus leaving a negligible impact on computation cost.

3.4 Loss Functions

We use a pixel-wise $L2$ loss for our model:

$$\min_{\Omega} \frac{1}{m \times n} \sum_{i,j} (S_{i,j} - D_{i,j})^2 \tag{3}$$

where Ω represents the trainable parameters set in SAU-Net and $S = F(I)$ in Equation 1, namely the output from the SAU-Net. The subscripts i, j denotes the pixel location and $1 <= i <= m$, $1 <= j <= n$. During training, the actual loss value can be very small due to the Gaussian filtering and this render the network difficult to train. We solve this by scaling the loss by a constant large value, e.g., 100, in training and dividing it during inference.

We also tried to add a loss function to explicitly minimize the cell counts but it only learned the artifacts in images. We think this is caused the ambiguity in the loss function since the loss function simply required the model to predict the overall cell counts, instead of detecting actual cells, and there are numerous mappings from the input image just to yield a correct cell count. A similar result was observed in [21].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Algorithm 2: A single layer of our online Batch Normalization

Input: A mini-batch B with size M of activations x: $B = \{x_1, \ldots, x_M\}$; trainable parameters: γ, β; constant for numerical stability: ϵ; test samples with size N: $T_{test} = \{x_1, \ldots, x_N\}$, train samples with size L: $T_{train} = \{x_1, \ldots, x_L\}$,

Training:

/* mini-batch mean and variance */

$\mu_B \leftarrow mean(\{x_1, \ldots, x_M\})$;

$\sigma_B \leftarrow std(\{x_1, \ldots, x_M\}) + \epsilon$;

/* normalize activations within the mini-batch */

for $i = 1 \ldots M$ **do**

  $y_i \leftarrow \gamma \frac{x_i - \mu_B}{\sigma_B} + \beta$

end

optimize γ, β via back-propagation ;

Inference:

for $j = 1 \ldots N$ **do**

  /* re-normalize */

  $\mu \leftarrow mean(\{x_1, \ldots, x_L, x_j\})$;

  $\sigma \leftarrow std(\{x_1, \ldots, x_L, x_j\}) + \epsilon$;

  $y_j \leftarrow \gamma \frac{x_j - \mu}{\sigma} + \beta$

end

4 DATASETS

We evaluate the propose method on four public benchmarks: synthetic fluorescence microscopy (VGG) dataset [13], Modified Bone Marrow (MBM) dataset [10], human subcutaneous adipose tissue (ADI) dataset [21] and Dublin Cell Counting (DCC) dataset [19]. A comparison of the datasets is listed in Table 1 and sample images from each dataset are shown in Figure 3.

- **VGG:** Lempitsky and Zisserman [13] used the method in [12] to create VGG dataset, which simulated bacterial cells from fluorescence-light microscopy at various focal distances.
- **MBM:** Cohen et al. [21] introduced the bone marrow Modified Bone Marrow (MBM) dataset based on the dataset first released by Kainz et al. [10]. This dataset contains real images of human bone marrow with various cell types stained in blue.
- **ADI:** Human subcutaneous adipose tissue (ADI) dataset [21] are constructed from the Genotype Tissue Expression Consortium [15] with densely packed adipocyte cells.
- **DCC:** Marsden et al. [19] built Dublin Cell Counting (DCC) dataset to represent a wide range of cells, including embryonic mice stem cells, human lung adenocarcinoma, human monocytes, etc. The image size ranges from 306 × 322 to 798 × 788, intended to increase the variation of the dataset.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

5 RESULTS

In this section, we first describe implementation details for training and then the configurations for the experiment, followed by quantitative results across multiple datasets. The implementation parameters mentioned in the section, except the batch size, are shared across all four datasets to further demonstrate the versatility of the proposed model. Based on the size of the image for each dataset, the batch size is set to 15 for MBM dataset, and 75 for the remaining three datasets.

5.1 Implementation Details

5.1.1 Optimization.—We choose Adam optimizer with decoupled weight decay [17], which explicitly enforces weight decay instead of implemented as L_2 regularization, to train our network. The weight decay for the Adam optimizer is set to 0.001. The weights of the network are randomly initialized by Glorot method [6] for every experiment. For the learn rate, we use a cosine annealing schedule with warm restarts [16], which incrementally lower the learning rate based on a cosine decay function. The initial value for the schedule is set to 0.001 and the restart step is set to 50 with a multiplier of 2. Each experiment is iterated for 350 steps. The actual learning rate is shown in Figure 4.

5.1.2 Prepossessing.—Our network can work with the raw image without sophisticated prepossessing. Given the varied image size in DCC dataset, we downsample the image and the density together to 256 × 256, and linearly scale the density map to ensure the same cell count. We also pad the image edge from 150 × 150 to 152 × 152 with zeros in ADI dataset so that the image size is a multiple of 8, convenient to max-pooling in the network. The images in the other two datasets remain the same.

5.1.3 Data Augmentation.—During training, we randomly crop 87.5% region of the images and apply random horizontal flipping, vertical flipping and rotate 90 degree rotation. Other non-90 degree rotation is not considered due to potential information loss during interpolation. Dropout layers are also added in the end of encoding path of SAU-Net to reduce over-fitting, as in Figure 2.

5.2 Experiment Configurations

We use Mean Absolute Error (MAE) of cell counts between prediction and ground-truth per image as the evaluation criteria. The training and testing splits in each dataset are randomly selected for each trial, and we repeat the experiment ten times, reporting the mean and variance of MAE. We first conduct an ablation study to examine the effectiveness of each proposed component, and then we compare our method with the state-of-the-art techniques in each dataset and list their best performance from their original work.

5.2.1 Ablation Study.—We conduct the ablation study on VGG dataset, favoring its popularity. We follow the same experiment configuration mentioned in Section 5.2 by randomly selecting 64 images out of 200 images for training and report the result in Table 2. It shows that the default batch normalization does not work well on small datasets when used in conjunction with data augmentation, as discussed in Section 3.3, and our online

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

batch normalization effectively solves this problem. It also demonstrates that the proposed network expansion with the Self-Attention module can further improve the performance of our model.

5.2.2 VGG.—In this synthetic dataset, our performance is on par with the leading methods. The result and a sample prediction are shown in Table 3 and Figure 5. Note that the leading models are highly engineered for this synthetic dataset and our method achieve the state-of-the-art performance in the rest real datasets.

5.2.3 MBM.—Our method outperforms other leading methods in this real dataset. Table 4 and Figure 6 show the result and a sample prediction.

5.2.4 ADI.—Similarly, the proposed method achieves the state-of-the-art performance in this dataset. Table 5 shows the result. As illustrated in Figure 7, this dataset represents a challenging scenario given the complicated cell structure.

5.2.5 DCC.—Table 6 also shows the superior result of our method in this recently released real dataset. A sample prediction is provided in Figure 8.

6 CONCLUSION

In this paper, we propose a novel deep network structure which employs a self-attention module and is capable of universally solving the cell counting task for different cell types. We also extend the current Batch Normalization method with a online version to better adapt to small datasets. We assess the proposed method on four public benchmarks and our model has consistently shown superior or competitive performance to the state-of-the-art methods.

Future work in this area includes jointly-trained models based on adversarial learning, which could take advantage of data from multiple sources and lead to further improvement.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation under grants OCI-1153775 and OAC-1649916.

REFERENCES

[1]. Arteta Carlos, Lempitsky Victor, Noble J. Alison, and Zisserman Andrew. 2012. Learning to Detect Cells Using Non-overlapping Extremal Regions. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2012, Ayache Nicholas, Delingette Hervé, Golland Polina, and Mori Kensaku (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 348–356.

[2]. Arteta Carlos, Lempitsky Victor, Noble J. Alison, and Zisserman Andrew. 2014. Interactive Object Counting. In Computer Vision – ECCV 2014, Fleet David, Pajdla Tomas, Schiele Bernt, and Tuytelaars Tinne (Eds.). Springer International Publishing, Cham, 504–518.

[3]. Arteta Carlos, Lempitsky Victor, J. Alison Noble, and Andrew Zisserman. 2016. Detecting overlapping instances in microscopy images using extremal region trees. Medical Image Analysis 27 (2016), 3–16. Discrete Graphical Models in Biomedical Image Analysis. [PubMed: 25980675]

[4]. Bernier Raphael, Golzio Christelle, Xiong Bo, Stessman Holly A, Coe Bradley P, Penn Osnat, Witherspoon Kali, Gerdts Jennifer, Baker Carl, Vulto-van Silfhout Anneke T, et al. 2014.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Disruptive CHD8 mutations define a subtype of autism early in development. Cell 158, 2 (2014), 263–276. [PubMed: 24998929]

[5]. Fiaschi L, Koethe U, Nair R, and Hamprecht FA. 2012. Learning to count with regression forest and structured labels. In Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). 2685–2688.

[6]. Glorot Xavier and Bengio Yoshua. 2010. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the thirteenth international conference on artificial intelligence and statistics. 249–256.

[7]. Guo Yue, Wang Qian, Krupa Oleh, Stein Jason, Wu Guorong, Bradford Kira, and Krishnamurthy Ashok. 2019. Cross Modality Microscopy Segmentation via Adversarial Adaptation. In Bioinformatics and Biomedical Engineering, Rojas Ignacio, Valenzuela Olga, Rojas Fernando, and Ortuño Francisco (Eds.). Springer International Publishing, Cham, 469–478.

[8]. Ioffe Sergey. 2017. Batch Renormalization: Towards Reducing Minibatch Dependence in Batch-Normalized Models. CoRR abs/1702.03275 (2017). arXiv:1702.03275

[9]. Ioffe Sergey and Szegedy Christian. 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. CoRR abs/1502.03167 (2015). arXiv:1502.03167

[10]. Kainz Philipp, Urschler Martin, Schulter Samuel, Wohlhart Paul, and Lepetit Vincent. 2015. You should use regression to detect cells. In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 276–283.

[11]. Krizhevsky Alex, Sutskever Ilya, and Hinton Geoffrey E. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In Advances in Neural Information Processing Systems 25, Pereira F, Burges CJC, Bottou L, and Weinberger KQ (Eds.). Curran Associates, Inc., 1097–1105.

[12]. Lehmussola Antti, Ruusuvuori Pekka, Selinummi Jyrki, Huttunen Heikki, and Yli-Harja Olli. 2007. Computational framework for simulating fluorescence microscope images with cell populations. IEEE transactions on medical imaging 26, 7 (2007), 1010–1016. [PubMed: 17649914]

[13]. Lempitsky Victor and Zisserman Andrew. 2010. Learning to count objects in images. In Advances in neural information processing systems. 1324–1332.

[14]. Long J, Shelhamer E, and Darrell T. 2015. Fully convolutional networks for semantic segmentation. In 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 3431–3440. 10.1109/CVPR.2015.7298965

[15]. Lonsdale John, Thomas Jeffrey, Salvatore Mike, Phillips Rebecca, Lo Edmund, Shad Saboor, Hasz Richard, Walters Gary, Garcia Fernando, Young Nancy, et al. 2013. The genotype-tissue expression (GTEx) project. Nature genetics 45, 6 (2013), 580. [PubMed: 23715323]

[16]. Loshchilov Ilya and Hutter Frank. 2016. Sgdr: Stochastic gradient descent with warm restarts. arXiv preprint arXiv:1608.03983 (2016).

[17]. Loshchilov Ilya and Hutter Frank. 2017. Fixing Weight Decay Regularization in Adam. CoRR abs/1711.05101 (2017). arXiv:1711.05101

[18]. Lu Erika, Xie Weidi, and Zisserman Andrew. 2018. Class-agnostic counting. arXiv preprint arXiv:1811.00472 (2018).

[19]. Marsden Mark, McGuinness Kevin, Little Suzanne, Keogh Ciara E, and O'Connor Noel E. 2018. People, Penguins and Petri Dishes: Adapting Object Counting Models To New Visual Domains And Object Types Without Forgetting. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 8070–8079.

[20]. Oktay Ozan, Schlemper Jo, Folgoc Loïc Le, Lee Matthew C. H., Heinrich Mattias P., Misawa Kazunari, Mori Kensaku, McDonagh Steven G., Hammerla Nils Y., Kainz Bernhard, Glocker Ben, and Rueckert Daniel. 2018. Attention U-Net: Learning Where to Look for the Pancreas. CoRR abs/1804.03999 (2018). arXiv:1804.03999

[21]. Cohen Joseph Paul, Boucher Genevieve, Glastonbury Craig A, Lo Henry Z, and Bengio Yoshua. 2017. Count-ception: Counting by fully convolutional redundant counting. In Proceedings of the IEEE International Conference on Computer Vision. 18–26.

[22]. Polley Mei-Yin C, Leung Samuel CY, McShane Lisa M, Gao Dongxia, Hugh Judith C, Mastropasqua Mauro G, Viale Giuseppe, Zabaglo Lila A, Penault-Llorca Frédérique, Bartlett

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

John MS, et al. 2013. An international Ki67 reproducibility study. Journal of the National Cancer Institute 105, 24 (2013), 1897–1906. [PubMed: 24203987]

[23]. Ren Shaoqing, He Kaiming, Girshick Ross, and Sun Jian. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In Advances in Neural Information Processing Systems 28, Cortes C, Lawrence ND, Lee DD, Sugiyama M, and Garnett R (Eds.). Curran Associates, Inc., 91–99.

[24]. Ronneberger Olaf, Fischer Philipp, and Brox Thomas. 2015. U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention. Springer, 234–241.

[25]. Vaswani Ashish, Shazeer Noam, Parmar Niki, Uszkoreit Jakob, Jones Llion, Gomez Aidan N., Kaiser Lukasz, and Polosukhin Illia. 2017. Attention Is All You Need. CoRR abs/1706.03762 (2017). arXiv:1706.03762

[26]. Wang Limin, Xiong Yuanjun, Wang Zhe, Qiao Yu, Lin Dahua, Tang Xiaoou, and Van Gool Luc. 2016. Temporal Segment Networks: Towards Good Practices for Deep Action Recognition. CoRR abs/1608.00859 (2016). arXiv:1608.00859

[27]. Wang Xiaolong, Girshick Ross, Gupta Abhinav, and He Kaiming. 2018. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 7794–7803.

[28]. Xie Weidi, Noble J Alison, and Zisserman Andrew. 2018. Microscopy cell counting and detection with fully convolutional regression networks. Computer methods in biomechanics and biomedical engineering: Imaging & Visualization 6, 3 (2018), 283–292.

[29]. Xue Yao, Ray Nilanjan, Hugh Judith, and Bigras Gilbert. 2016. Cell counting by regression using convolutional neural network. In European Conference on Computer Vision. Springer, 274–290.

[30]. Guo Yue, Wrammert J, Singh K, KC A, Bradford K, and Krishnamurthy A. 2016. Automatic analysis of neonatal video data to evaluate resuscitation performance. In 2016 IEEE 6th International Conference on Computational Advances in Bio and Medical Sciences (ICCABS). 1–6. 10.1109/ICCABS.2016.7802775

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

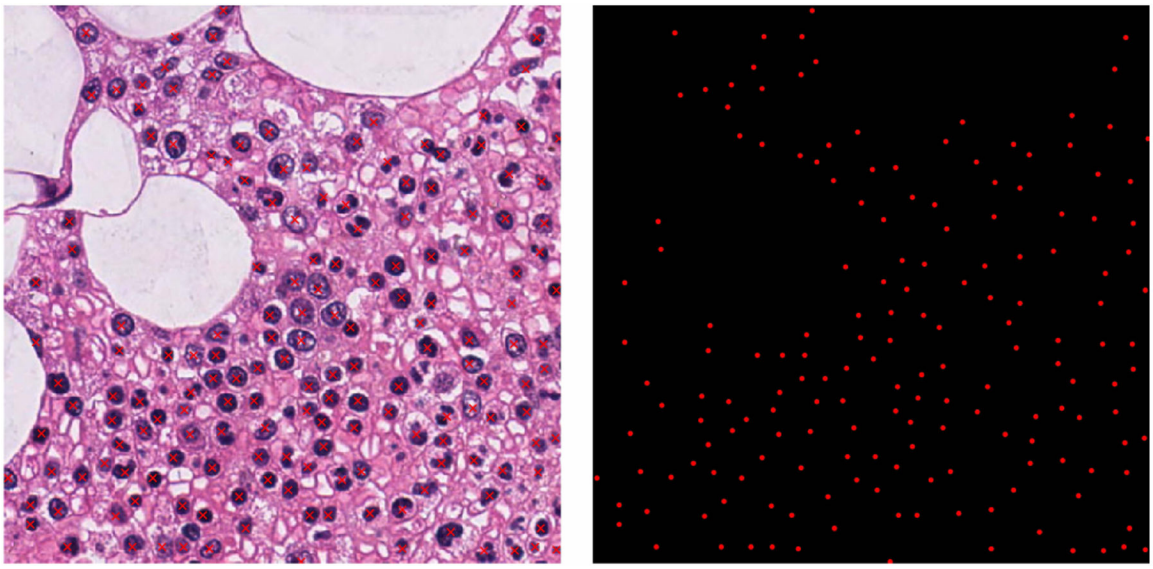

Figure 1:
A sample image from Modified Bone Marrow (MBM) dataset [10] with dot annotations shown as red cross overlays (left image) and the corresponding dot annotations used in training (right image).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

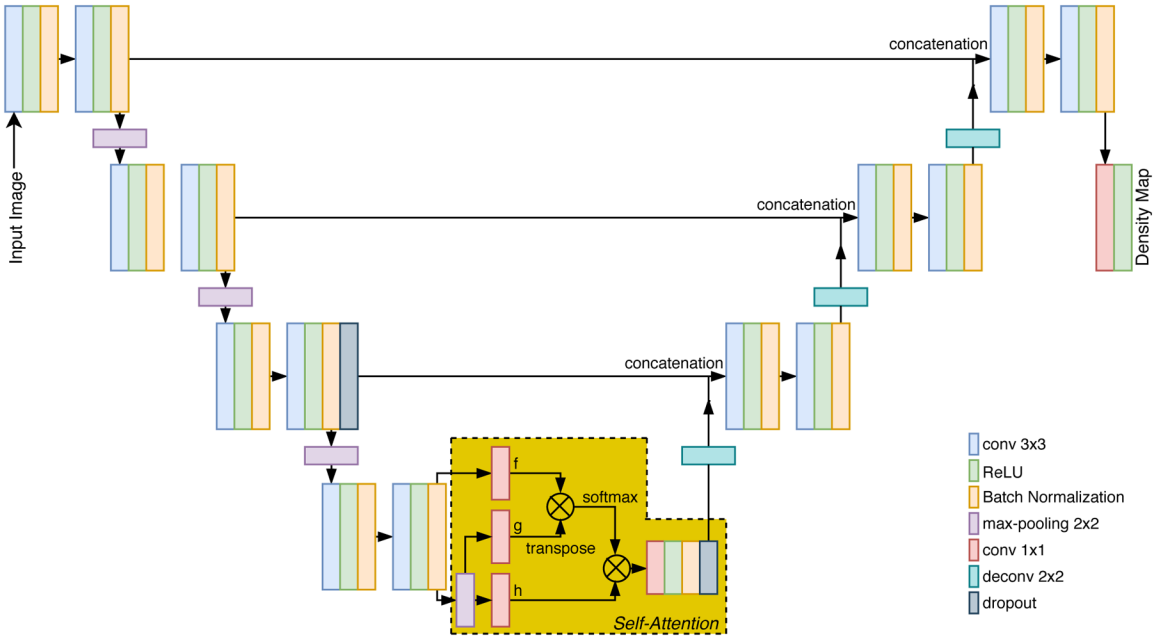

Figure 2:

The overall structure of SAU-Net. SAU-Net is essentially a U-Net with the addition of a Self-Attention module, and modified batch normalization. In the Self-Attention module, f, g and h are linear embeddings, implemented as 1×1 convolution with proper reshaping. ⊗ denotes matrix multiplication.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

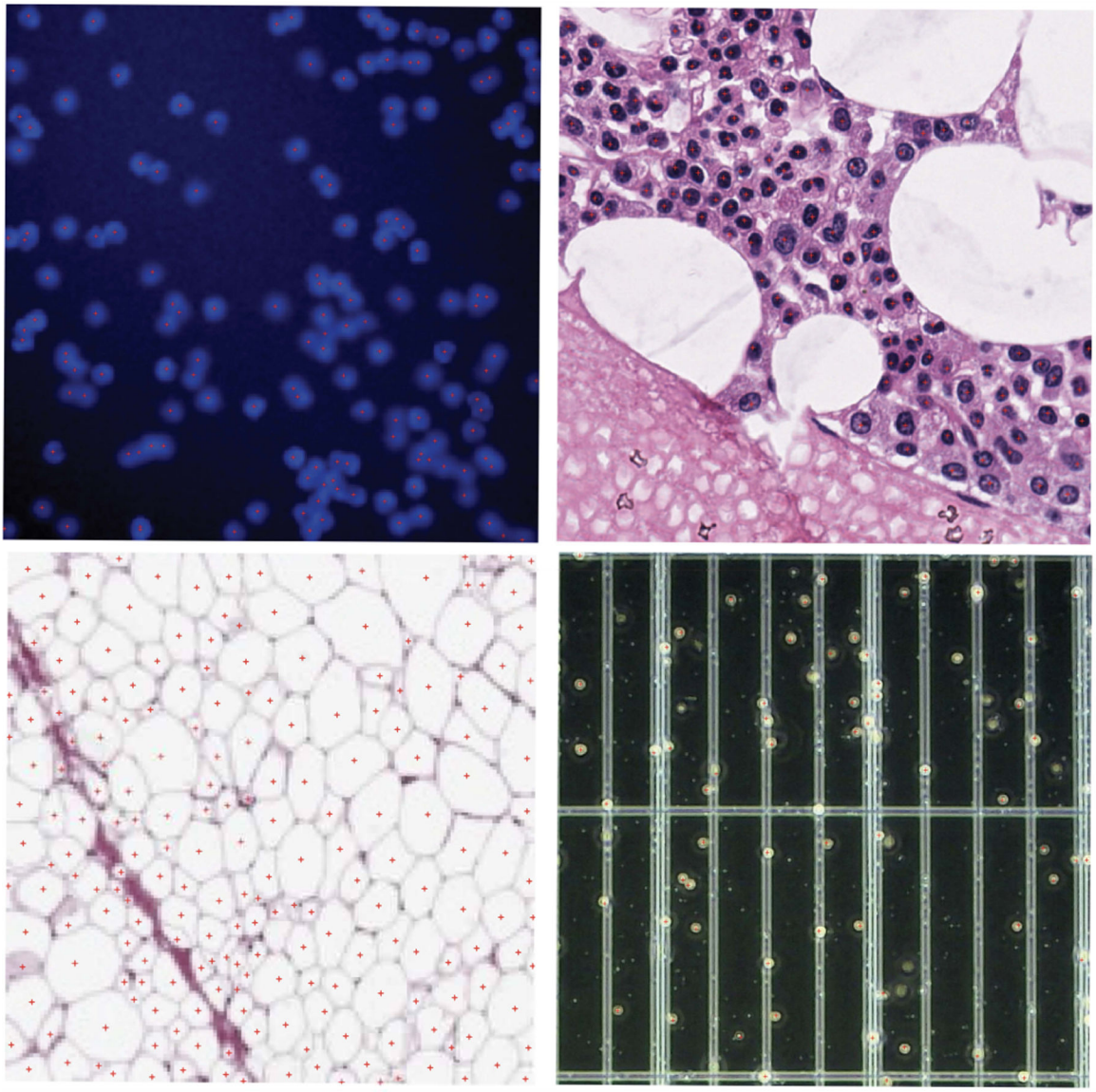

Figure 3:
Sample images with dot annotations as the red cross overlays from four public benchmark datasets: VGG (top left), MBM (top right), ADI (bottom left), DCC (bottom right).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

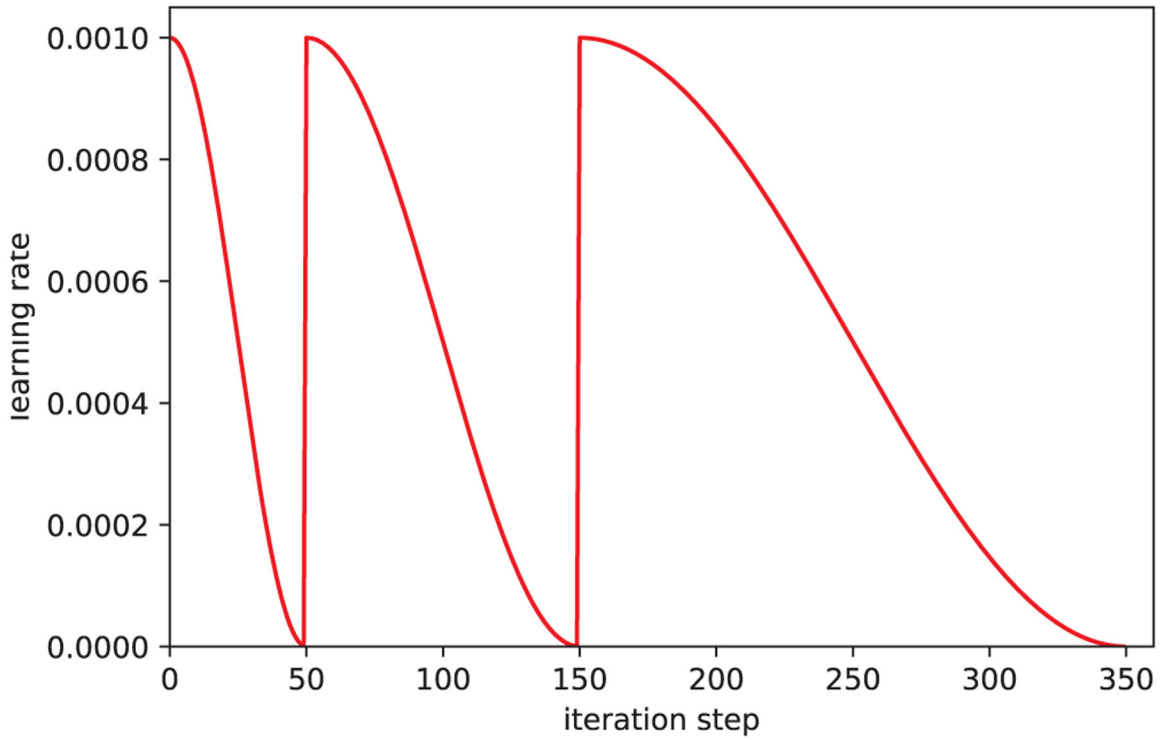

Figure 4:

The cosine annealing learning rate schedule with warm restarts. The initial value is 0.001 and the restart step is set to 50 with a multiplier of 2. The total iteration step is 350 with 2 restarts at Step 50 and Step 150.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

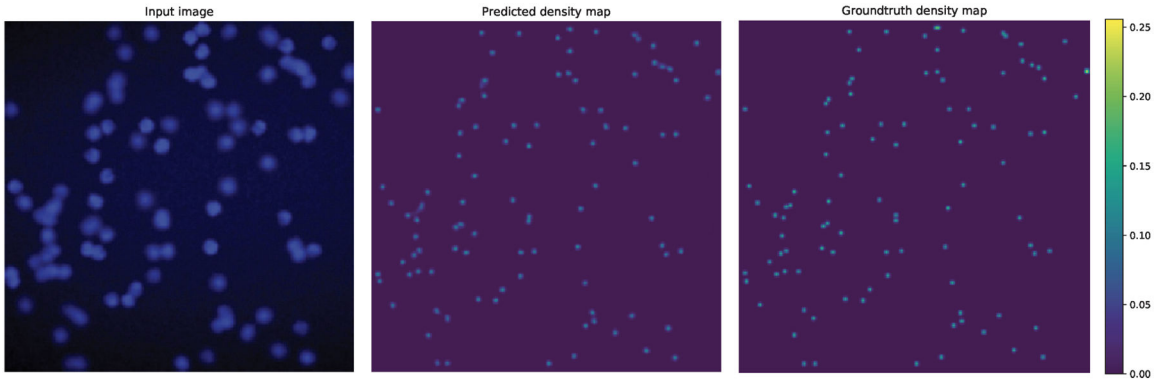

Figure 5:
Sample predicted density map on the test set for VGG dataset. Groundtruth Cell Count: 100, Predicted: 100.9

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

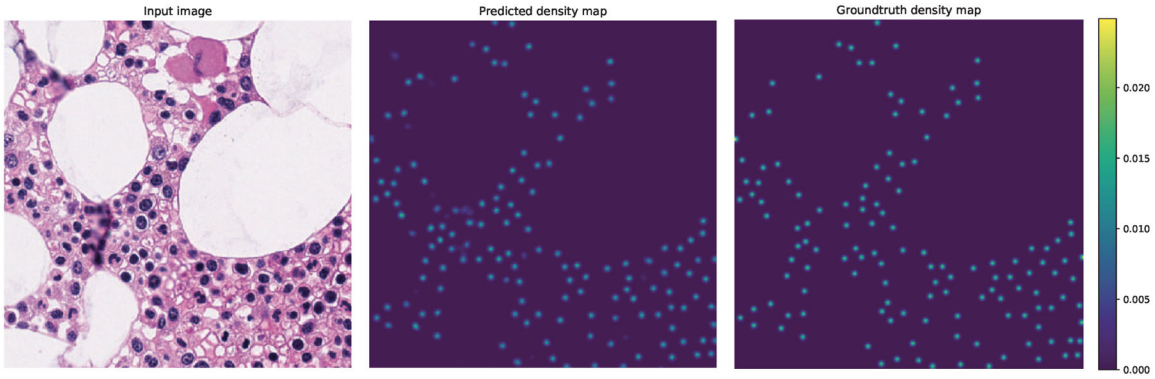

Figure 6:
Sample predicted density map on the test set for MBM dataset. Groundtruth Cell Count: 134, Predicted: 135.8

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

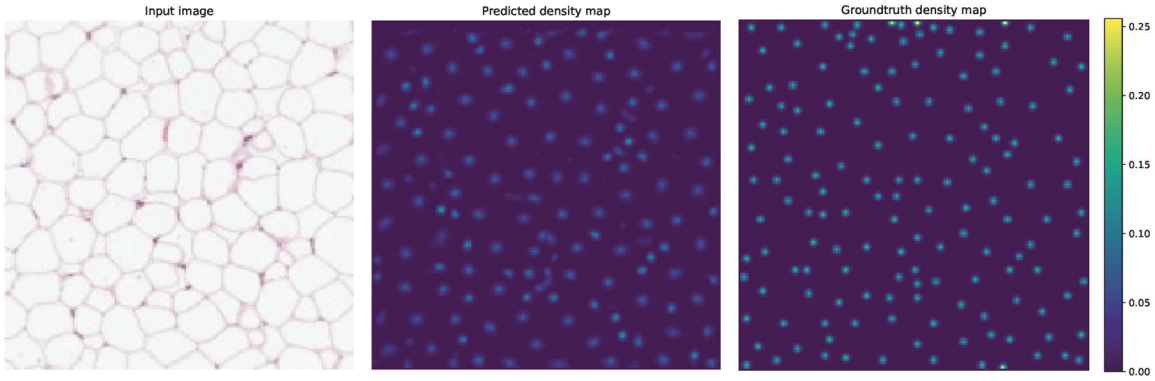

Figure 7:
Sample predicted density map on the test set for ADI dataset. Groundtruth Cell Count: 149, Predicted: 142.1

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

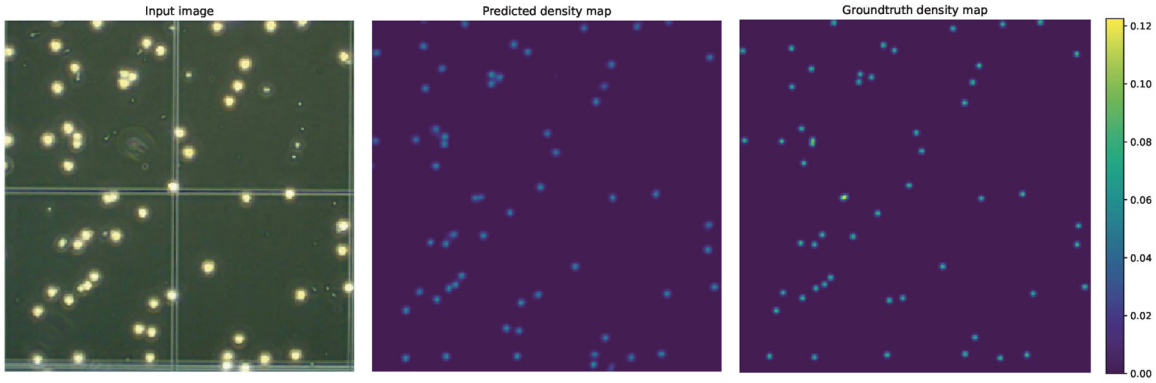

Figure 8:
Sample predicted density map on the test set for DCC dataset. Groundtruth Cell Count: 55, Predicted: 58.9

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1:

A comparison of four datasets. N_{train} and N_{total} are the training image number and total image number, respectively. N_{train} is chosen from common settings for comparison convenience with other methods. Cell Count denotes the mean cell count per image and the corresponding variance over all the images.

Dataset Name	Image Size	N_{train} /N_{total}	Cell Count	Type
VGG [13]	256 × 256	64/200	174 ± 64	synthetic
MBM [10]	600 × 600	15/44	126 ± 33	real
ADI [21]	150 × 150	50/200	165 ± 44	real
DCC [19]	varied	100/176	34 ± 22	real

ACM BCB. Author manuscript; available in PMC 2021 May 26.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Ablation study on VGG

Method	MAE
U-Net + Batch Normalization	45.4 ± 15.0
U-Net + Online Batch Normalization	2.9 ± 0.7
U-Net + Online Batch Normalization + Self-Attention (proposed)	2.6 ± 0.4

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3:

VGG (200 images in total)

Method	MAE	N_{train}
ResNet-152 (R), Xue et al. [29]	7.5 ± 2.2	100
GMN, Lu et al. [18]	3.6 ± 0.3	32
FCRN-A, Xie et al. [28]	2.9 ± 0.2	64
Count-Ception, Cohe el al. [21]	**2.3 ± 0.4**	50
SAU-Net (proposed)	2.6 ± 0.4	64

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4:

MBM (44 images in total)

Method	MAE	N_{train}
FCRN-A, Xie et al. [28]	21.3 ± 9.4 *	15
Marsden el al. [19]	20.5 ± 3.5	15
Count-Ception, Cohe el al. [21]	8.8 ± 2.3	15
SAU-Net (proposed)	**5.7 ± 1.2**	15

* Implemented by Cohe el al. [21].

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5:

ADI (200 images in total)

Method	MAE	N_{train}
Count-Ception, Cohe el al. [21]	19.4 ± 2.2	50
SAU-Net (proposed)	**14.2 ± 1.6**	50

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 6:

DCC (176 images in total)

Method	MAE	N_{train}
Marsden el al. [19]	8.4 *	100
SAU-Net (proposed)	**3.0 ± 0.3**	100

* Reported in a fixed split in their work.