Preliminary Results on Eye Diameter Prediction

John Lipor, Andrew Franck, Hadi Khazaei, and Faryar Etesami

July 27, 2023

1 Introduction

We consider a dataset consisting of 107 ultrasound images of eyes captured with the Butterfly iQ+ ultrasound probe. Of these images, 77 are taken from typical/healthy patients, while 30 are taken from eyes with known abnormalities. The goal is to utilize a convolutional neural network (CNN) to predict both the vertical and horizontal diameter of each eye, removing the need for the costly procedure of expert analysis of images.

2 Methodology

The dataset is split into a training and validation set as follows. For the typical images, we perform an 80/20 split, so that 64 images are used for training and 13 are used for validation. For the atypical images, we allot 20 to the training set and 10 to the test set, in order to ensure that there sufficiently many atypical images in the validation set. The training and validation datasets are much smaller than those typically used when training a CNN. As a result, we investigated the use of various forms of image augmentation. However, we did not find any augmentation to be beneficial to the training process. We note that, in contrast to images of typical objects (e.g., dogs, cars), ultrasound images have distinct spatial structure, with the main image having a specific orientation and an objective scale on the right side. This fact may account for the lack of benefit seen, and further investigation with larger labeled datasets is a topic of future research.

Our model is based on the Network-in-Network (NiN) structure [1], which utilizes 1×1 convolutions to aggregate information across image channels without destroying spatial structure. While we experimented with other network types, we found that any network employing linear layers ultimately learned the median value of the dataset. However, we note that the choice of network structure may vary as more images are considered in the training set. Each NiN block consists of (1) a convolutional layer of arbitrary kernel size followed by a rectified linear unit (ReLU) (2) a 1×1 convolutional layer followed by a ReLU, and (3) a second 1×1 convolutional layer followed by batch normalization and a ReLU. The first convolutional layer of each NiN block learns filters of user-specified size to detect the salient features in the image, while the subsequent two 1×1 convolutional layers act as nonlinear predictions for each pixel across all image channels. In this way, the NiN block performs convolution and a nonlinear transformation without destroying spatial structure. For our model, we use four NiN blocks with 96 output channels (filters) each. The first layer uses a kernel size of 11×11, while the next three use 5×5 kernels. A final NiN block consists of two output channels using kernel size 3×3, and the final prediction is performed by a global average pooling over each channel. The first channel corresponds to prediction of vertical diameters, while the second corresponds to horizontal diameters.

We train the above network using the mean absolute error (MAE), optimized using Adam [2] with a learning rate of 0.01 over 1000 epochs.

3 Results

Fig. 1 shows the training and validation error versus epoch. Interestingly, even for a very large number of epochs, the validation loss does not appear to diverge greatly from the training loss. This is likely due to the homogeneity of the training and validation sets, in which the images are very similar. We note that a more careful consideration of overfitting may be necessary when larger training and validation sets become

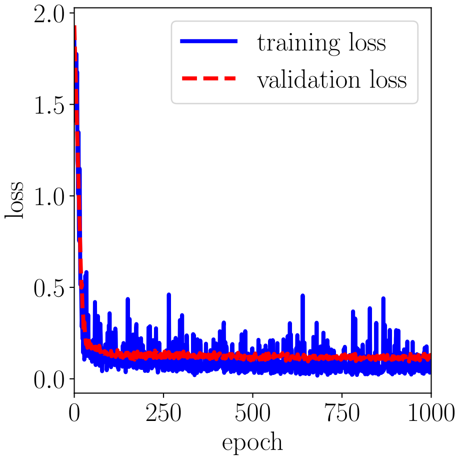
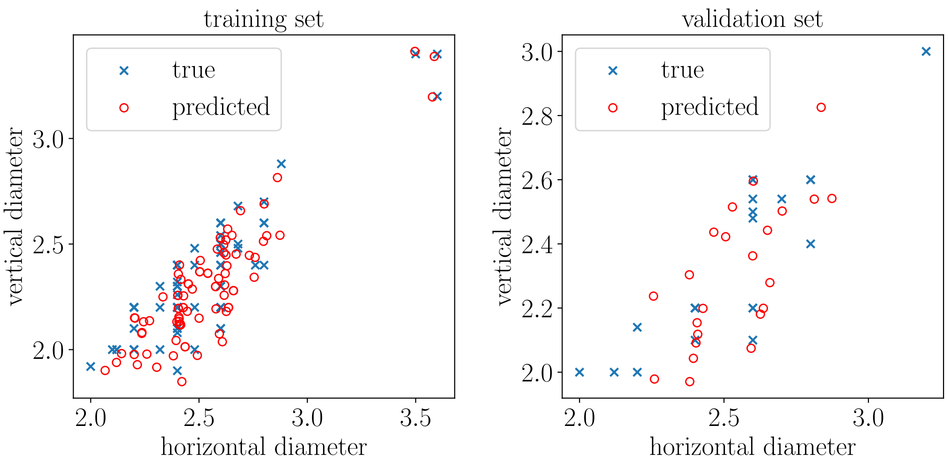


Figure 1: Training and validation loss for proposed NiN-based network.

available. The final validation MAE is 0.0925, corresponding to a relative error of 3.96%. For comparison, a simple algorithm that always predicts the median diameters of the training set results in a validation error of 8.72%, indicating that our trained network learns nontrivial predictions corresponding to the actual images, even from this small dataset. Fig.2 shows a scatter plot of the true and predicted diameters for (a) the training dataset and (b) the validation dataset. The figure shows that the network is able to accurately capture the training data while still generalizing to the validation set. For the validation data, we see that the network overpredicts the smallest horizontal diameters (bottom left points in Fig. 2(b)). These points likely correspond to images of atypical scenarios, for which there is less training data. Fig. 3 shows the images corresponding to the four most accurate predictions. In all cases, the entire eyeball is within the frame and largely surrounded by contrasting tissue, with the lens clearly visible. Fig. 4 shows the images corresponding to the worst four predictions. In this case, we see that images either have cloudiness within the eyeball region or dark regions that make it difficult to clearly delineate the boundary of the eyeball. We note that the worst two predictions correspond to the largest and smallest examples in the dataset, indicating that the algorithm is biased toward typical eyeball sizes. In the most extreme example (Fig. 4, right image), we see that the ultrasound scan covers much more of the image than other examples in the dataset, indicating that the model had a difficult time determining the absolute scale of the eye.

To gain insight into our network, we visualize the learned representations (*feature maps*) at the intermediate and output layers of the network. Fig. 5 shows example feature maps at the output of the first four NiN blocks for the most accurately predicted image (Fig. 3, top left). We see that the first NiN block performs broad-scale feature learning, picking out edges of various orientations and discovering contrasting regions. The second block appears to perform some smoothing, reducing the variation in locations away from the eyeball. The third block outlines both the eyeball and the ultrasound field, smoothing all other regions. Finally, the fourth block further refines and smooths these estimates. Fig. 6 shows the same outputs for the lowest-accuracy prediction (Fig. 4, bottom right). We note that the true vertical and horizontal diameters for this image are 20 mm and 20 mm, while the predicted diameters are 25 mm and 27 mm. The feature maps show that the model fails to determine the outline of the eyeball, and perhaps mistakes the outline of the ultrasound field for the eyeball in Fig. 6(d), which would explain the upward bias in the prediction for this image.

Finally, we consider a network saliency map obtained by guided backpropagation [3], which aims to determine which parts of the input image had the greatest impact on the final prediction by examining the gradients corresponding to each pixel location. For visualization purposes, we normalize all values and set values greater than 0.01 to 0.5; hence the resulting heatmaps do not indicate the strength of influence of each pixel. Fig. 7 shows the saliency maps for the four most accurate images, while Fig. 8 shows the maps for the worst four predictions. Red locations correspond to pixels that have greater influence over the model predictions. For the top four images, as well as the best two of the worst four predictions, we see



(a)

(b)

Figure 2: Scatter plots of true and predicted diameters for (a) training and (b) validation data. The results on validation data show the algorithm performs worst on outlying points with the smallest and largest diameters.

that the network focuses on bounding the ultrasound field as well as on a few extremal points within the eyeball itself. For the worst two predictions (Fig. 8, bottom row), we see that the model fails to determine the boundaries of either the eyeball or of the ultrasound field. Overall, we see that the network seeks to determine bounds around the region of interest within the image, which may provide a global scale, as well as a few points within the eyeball, which provide the relative size. Together, these result in the successful predictions reported above.

References

[1] M. Lin, Q. Chen, and S. Yan, "Network in network," *arXiv preprint arXiv:1312.4400*, 2013.

[2] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[3] J. T. Springenberg, A. Dosovitskiy, T. Brox, and M. Riedmiller, "Striving for simplicity: The all convolutional net," *arXiv preprint arXiv:1412.6806*, 2014.

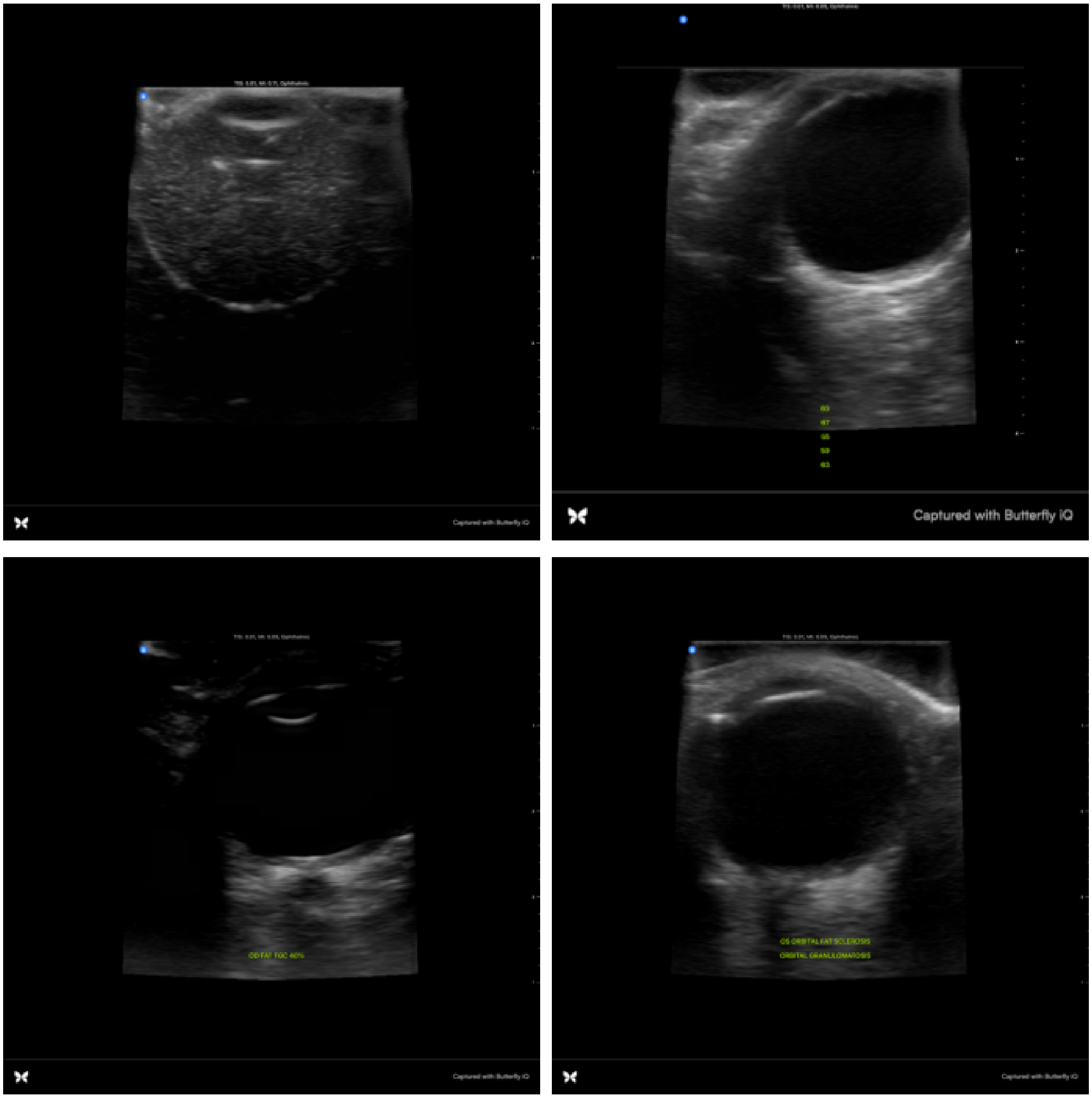

Captured with Butterfly iQ
Captured with Butterfly iQ
OD FAT TGC 40%
Captured with Butterfly iQ
OS ORBITAL FAT SCLEROSIS
ORBITAL GRANULOMATOSIS
Captured with Butterfly iQ

Figure 3: Ultrasound images corresponding to four most accurate predictions in validation set. Relative errors range from 0.12% - 0.71%. The images all contain eye images with clearly identified boundaries.

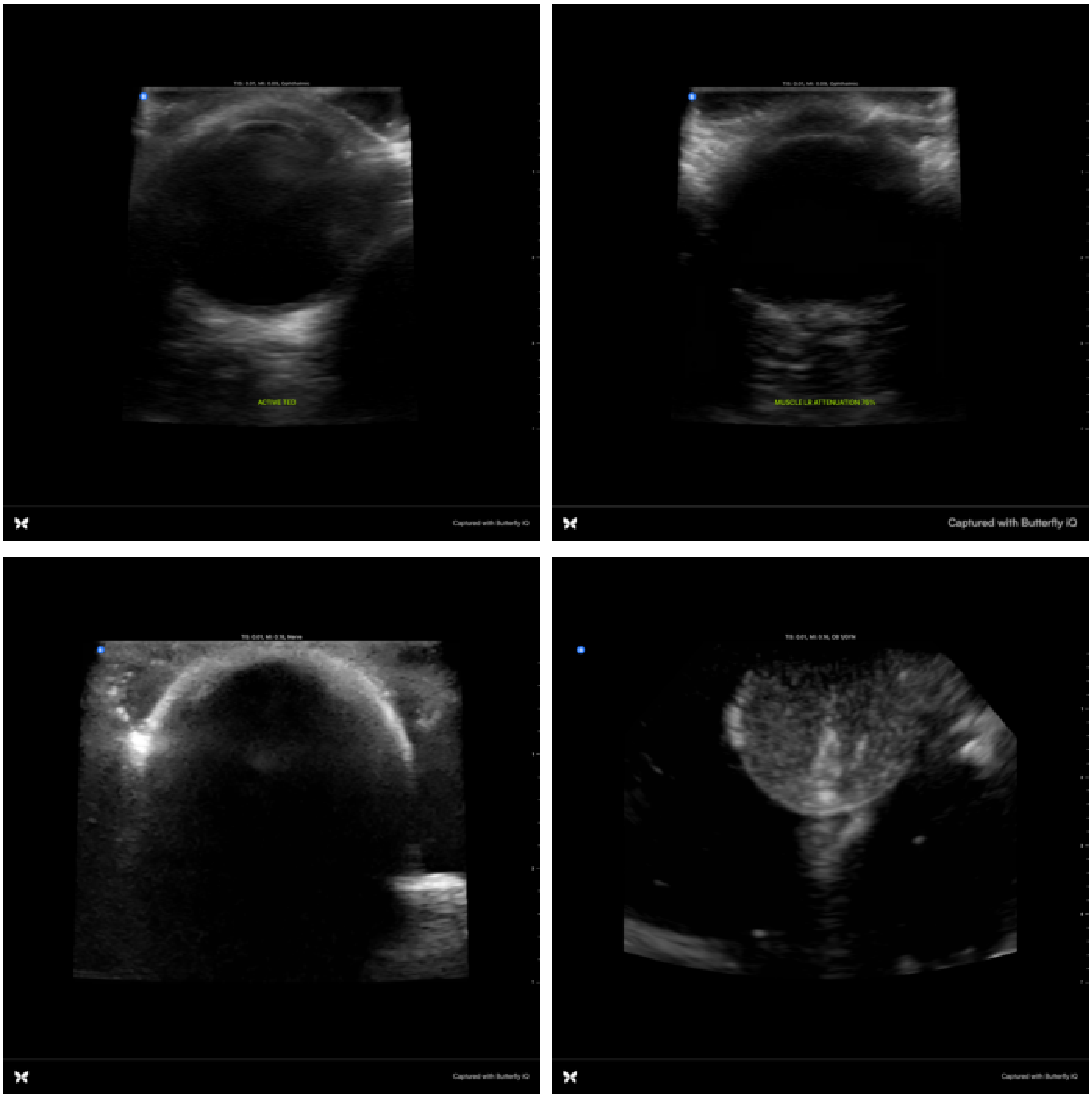
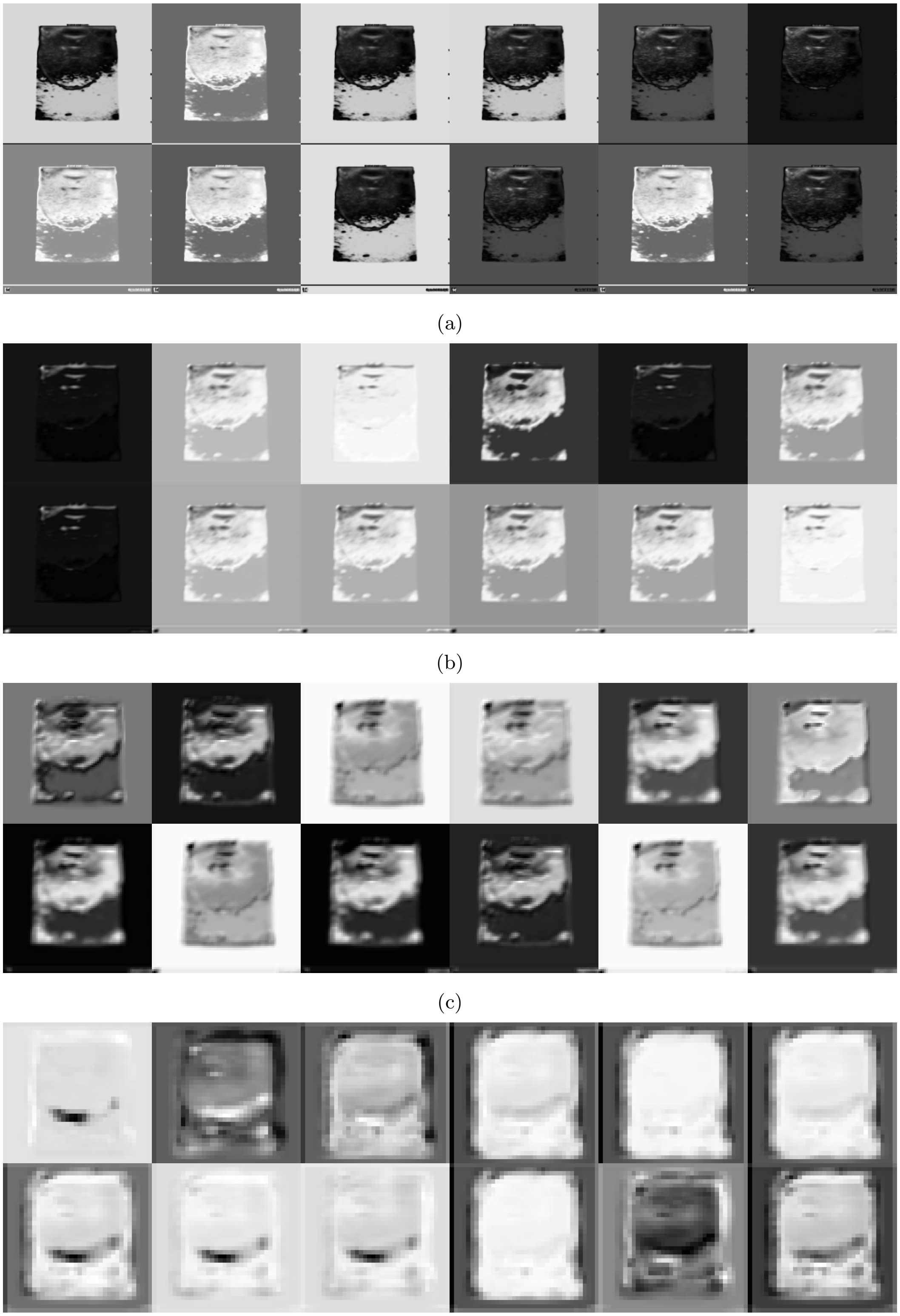

ACTIVE TED
Captured with Butterfly iQ
MUSCLE LR ATTENUATION 70%
Captured with Butterfly iQ
Captured with Butterfly iQ
Captured with Butterfly iQ

Figure 4: Ultrasound images corresponding to four least accurate predictions in validation set. Relative errors are 5.03%, 6.90%, 8.59%, and 30.09% (left to right, top to bottom). Failed predictions are likely due to lack of clear visual boundaries (first three images) or a different ultrasound field (fourth image).



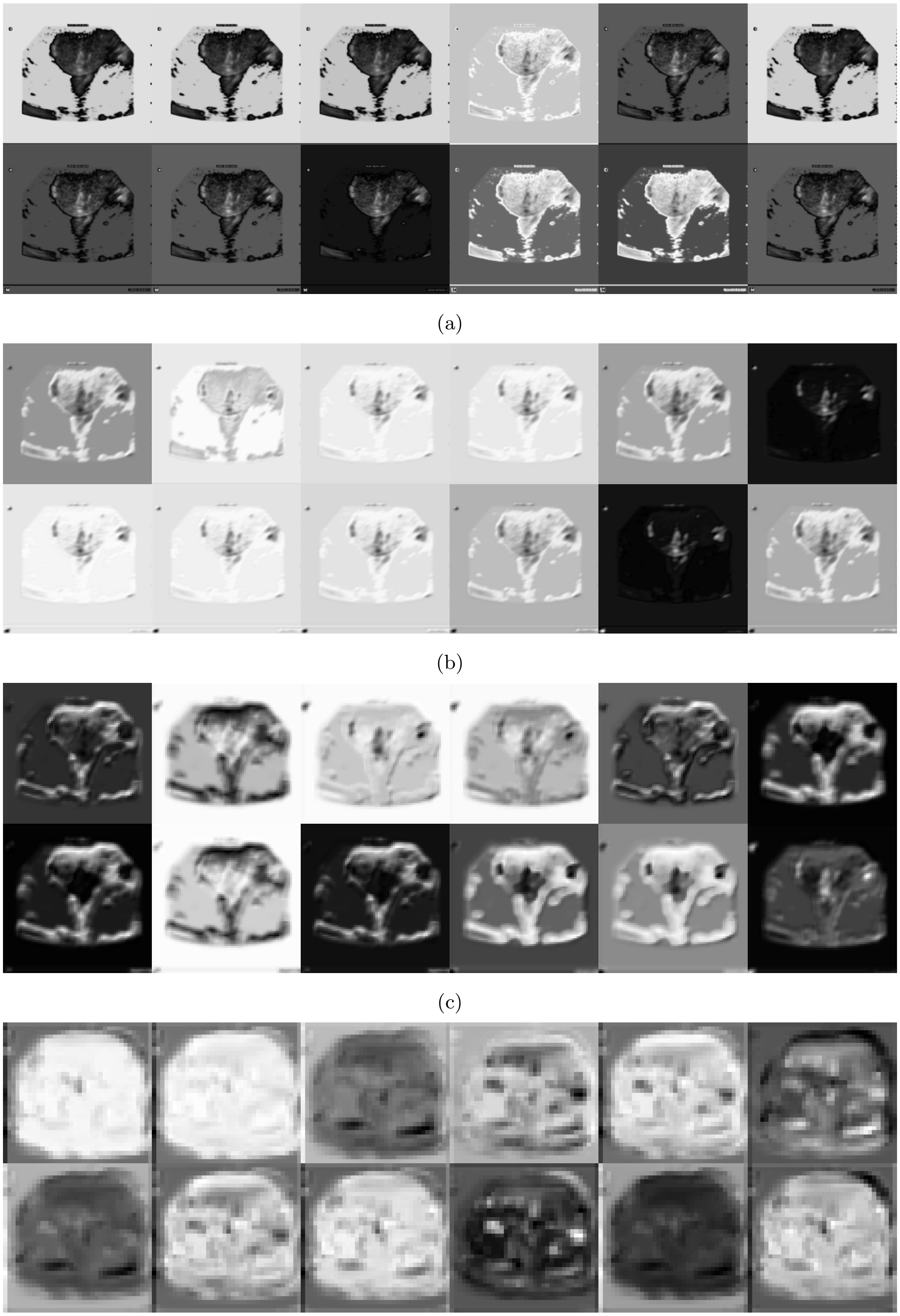
(a)

(b)

(c)

(d)

Figure 5: Feature maps output by each NiN block for lowest-error image. (a) First NiN block, (b) second block, (c) third block, (d) fourth block. At the fourth layer, the network appears to learn the boundary of the ultrasound field, along with an outline of the eye.



(a)

(b)

(c)

(d)

Figure 6: Feature maps output by each NiN block for highest-error image. (a) First NiN block, (b) second block, (c) third block, (d) fourth block. The third and fourth layer outputs indicate the model may have mistaken the ultrasound field for the eye, resulting in the high prediction for this image.

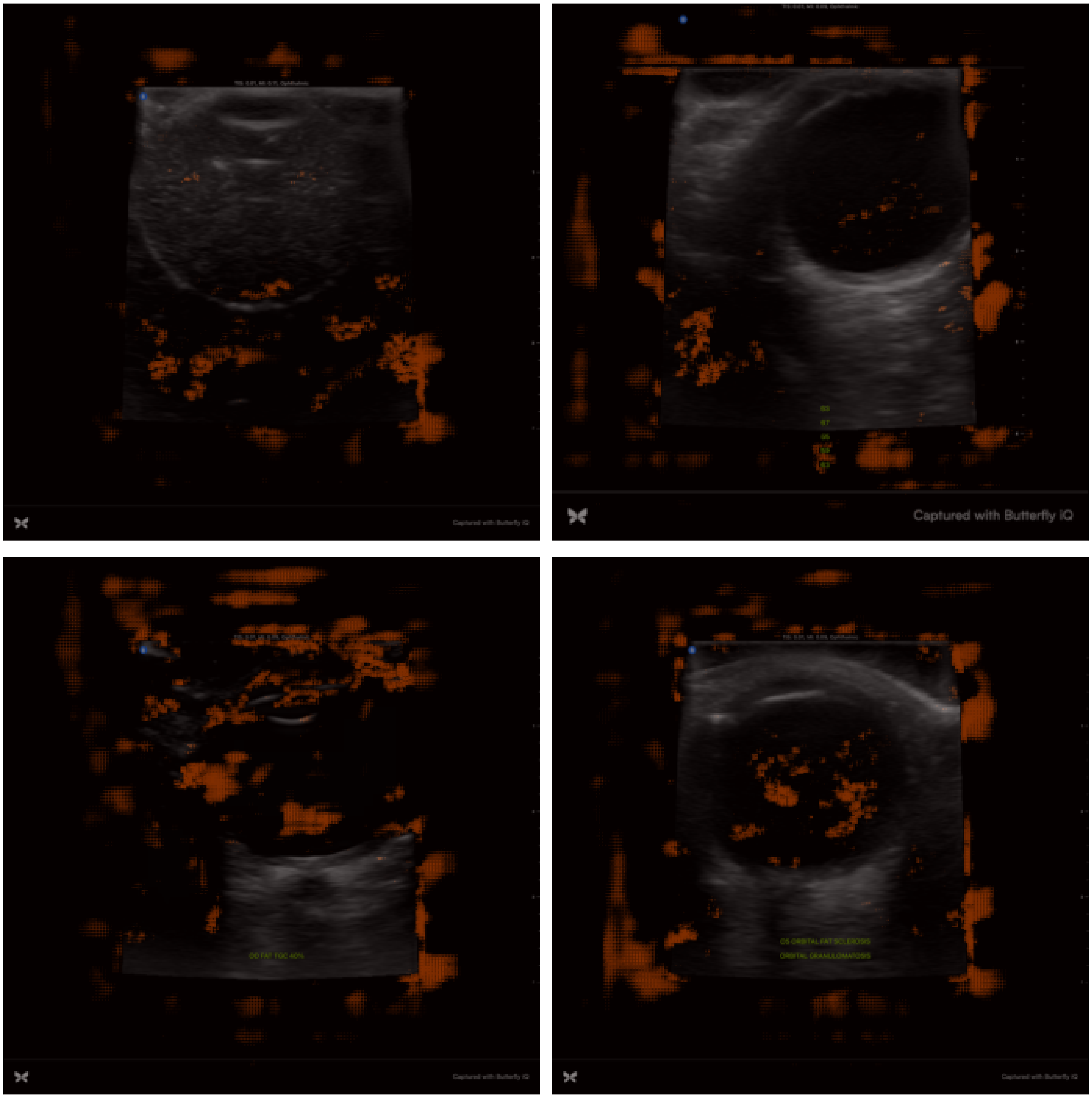


Figure 7: Saliency maps (via guided backpropagation) corresponding to four most accurate predictions in validation set. The model appears to determine the bounds of the ultrasound field, as well as the relative size of the eyeball within these bounds.

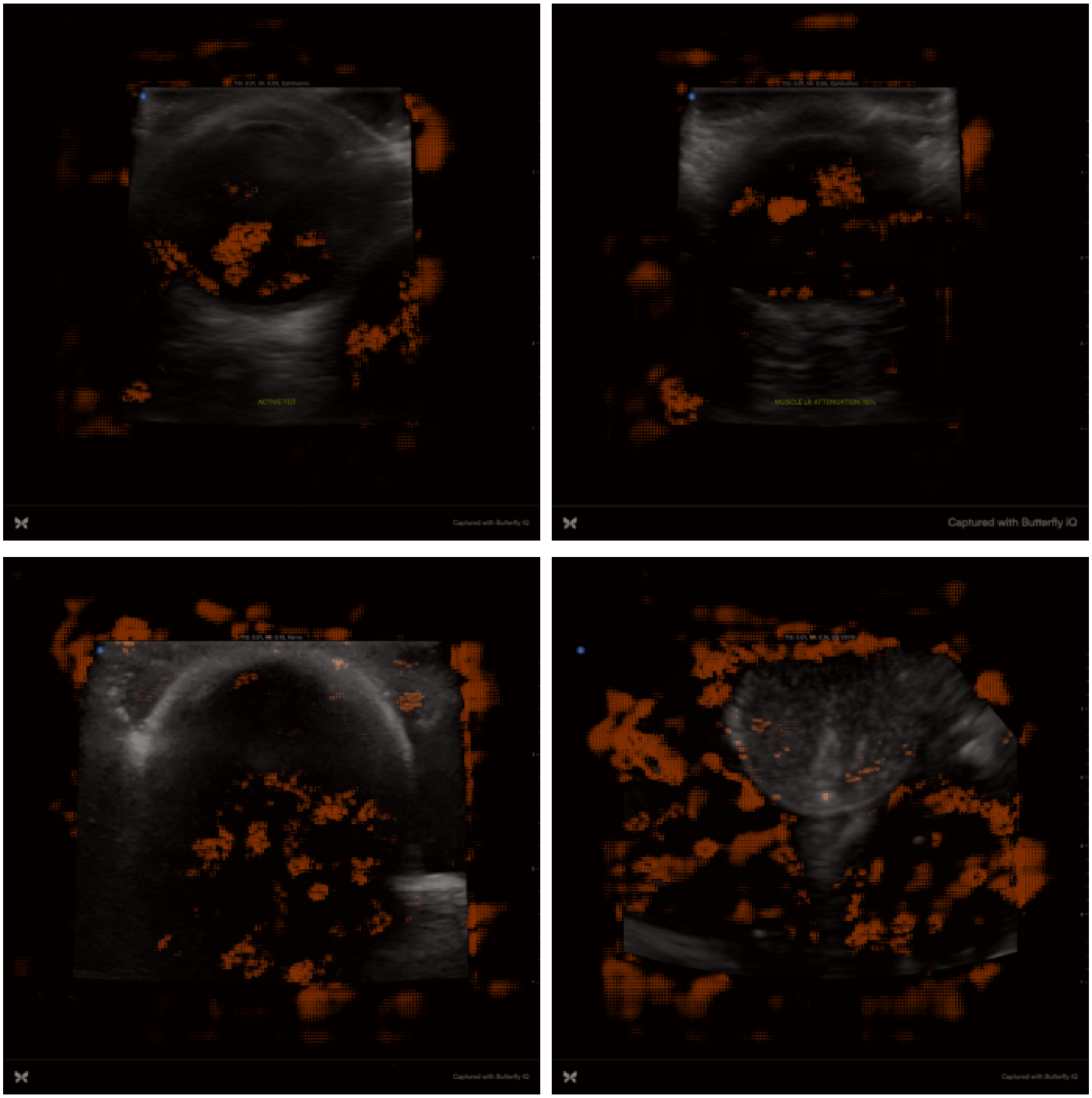

ACTIVE TED
Captured with Butterfly iQ
MUSCLE LR ATTENUATION 70%
Captured with Butterfly iQ
Captured with Butterfly iQ
Captured with Butterfly iQ

Figure 8: Saliency maps (via guided backpropagation) corresponding to four least accurate predictions in validation set. The model appears to either have trouble determining extremal points of the eyeball (bottom left) or bounds on the area of interest within the image (bottom right).