Team Information

Fanjie Gao, Yuyan Li

Explain the difference between the task of classification and segmentation, explain why there might be conflicts between the two tasks. (10 points)

Classification aims to label the entire image, while segmentation aims to label each pixel in the image. Classification only focuses on what is present in the image, while segmentation also captures the spatial information of objects in the image.

Introduce how FCN addresses the conflicts. Then introduce different versions of FCN, and explain how they balance the trade-off. (10 points)

FCN addresses the issue by implementing a fully convolutional network that can take in arbitrary-sized input and generate the same-sized output.

FCN-32s:
The output is very coarse, "The 32-pixel stride at the final prediction layer limits the scale of detail in the upsampled output" (p3436).

FCN-16s:
It adds a 2x upsampling layer of the final 32-pixel stride layer and combines the prediction with the intermediate layer. Then, the 16-pixel stride predictions are unsampled back to the image. It is able to make finer predictions based on the last, coarser layer.
FCN-8s:
It adds a 2x upsampling layer of the previous 16 stride layer and combines the prediction with an earlier intermediate layer. The rest is the same, the 8 pixel stride prediction is unsampled back. It has a diminishing return compared to FCN-16s.

Compare the evaluation metrics of pixel accuracy and IU introduced in the paper. Also compare mean IU and frequency-weighted IU. (10 points)

Pixel accuracy:
Percentage of correctly predicted pixels.

Mean accuracy
An average accuracy of every class.

Mean IU:
Average Intersection over Union of every class

Frequency-weighted IU:
IU score weighted with the frequency of the class

Pixel Accuracy vs. Intersection over Union (IU)
Pixel Accuracy offers a straightforward, albeit sometimes superficial, evaluation of segmentation accuracy. IU provides a more nuanced assessment by considering both false positives and false negatives, offering insight into the model's precision and recall for each class.

Mean IU vs. Frequency-weighted IU
Mean IU ensures all classes contribute equally to the evaluation, highlighting the model's ability to handle class imbalance. Frequency-weighted IU emphasizes the model's performance on the most common classes, making it useful for evaluating practical effectiveness in unbalanced datasets.

Comment on the limitations of FCN and potential rough directions for further improvements. (10 extra credits)

Limitations of FCNs
FCNs face challenges in processing contextual information and preserving image resolution, leading to potential misclassifications and blurred segmentation boundaries. They often struggle with class imbalance and scale variance, where the model biases towards frequent classes and fails to accurately segment objects of varying sizes due to a fixed receptive field size.

Directions for Improvements
Improving FCNs involves incorporating mechanisms for better global context understanding, such as attention mechanisms or graph convolutional networks, and developing sophisticated upsampling techniques to preserve spatial details. Addressing class imbalance through modified loss functions or data augmentation, and integrating scale-aware architectures or multi-modal data, can enhance segmentation accuracy. Additionally, including refinement modules like Conditional Random Fields (CRFs) for post-processing can further sharpen segmentation boundaries, offering a holistic improvement in FCN performance.

Training logs:

Epoch [1/50], Step [10/24], Loss: 3.8252
Epoch [1/50], Step [20/24], Loss: 1.9497
Pixel accuracy: 0.5979, Mean IoU: 0.0651, Frequency weighted IoU: 0.4131, Loss: 1.3934
Epoch [2/50], Step [10/24], Loss: 1.3182
Epoch [2/50], Step [20/24], Loss: 1.0135
Pixel accuracy: 0.6936, Mean IoU: 0.0894, Frequency weighted IoU: 0.5222, Loss: 1.1111
Epoch [3/50], Step [10/24], Loss: 0.9473
Epoch [3/50], Step [20/24], Loss: 0.8263
Pixel accuracy: 0.7589, Mean IoU: 0.1232, Frequency weighted IoU: 0.6211, Loss: 0.8727
Epoch [4/50], Step [10/24], Loss: 0.7645
Epoch [4/50], Step [20/24], Loss: 0.7605
Pixel accuracy: 0.7804, Mean IoU: 0.1386, Frequency weighted IoU: 0.6460, Loss: 0.8069
Epoch [5/50], Step [10/24], Loss: 0.7542
Epoch [5/50], Step [20/24], Loss: 0.6455
Pixel accuracy: 0.7717, Mean IoU: 0.1562, Frequency weighted IoU: 0.6501, Loss: 0.8954
Epoch [6/50], Step [10/24], Loss: 0.8988
Epoch [6/50], Step [20/24], Loss: 0.7309
Pixel accuracy: 0.7952, Mean IoU: 0.1539, Frequency weighted IoU: 0.6674, Loss: 0.7465
Epoch [7/50], Step [10/24], Loss: 0.6976
Epoch [7/50], Step [20/24], Loss: 0.6218
Pixel accuracy: 0.8185, Mean IoU: 0.1791, Frequency weighted IoU: 0.6951, Loss: 0.6464
Epoch [8/50], Step [10/24], Loss: 0.6192
Epoch [8/50], Step [20/24], Loss: 0.6067
Pixel accuracy: 0.8260, Mean IoU: 0.1900, Frequency weighted IoU: 0.7082, Loss: 0.6263
Epoch [9/50], Step [10/24], Loss: 0.5919
Epoch [9/50], Step [20/24], Loss: 0.5751
Pixel accuracy: 0.8261, Mean IoU: 0.1991, Frequency weighted IoU: 0.7173, Loss: 0.6172
Epoch [10/50], Step [10/24], Loss: 0.5387
Epoch [10/50], Step [20/24], Loss: 0.4995
Pixel accuracy: 0.8379, Mean IoU: 0.2224, Frequency weighted IoU: 0.7309, Loss: 0.5649
Epoch [11/50], Step [10/24], Loss: 0.4672
Epoch [11/50], Step [20/24], Loss: 0.4804
Pixel accuracy: 0.8173, Mean IoU: 0.2194, Frequency weighted IoU: 0.7171, Loss: 0.6692
Epoch [12/50], Step [10/24], Loss: 0.5739
Epoch [12/50], Step [20/24], Loss: 0.5586
Pixel accuracy: 0.8370, Mean IoU: 0.2214, Frequency weighted IoU: 0.7223, Loss: 0.5835
Epoch [13/50], Step [10/24], Loss: 0.4767
Epoch [13/50], Step [20/24], Loss: 0.4818
Pixel accuracy: 0.8492, Mean IoU: 0.2433, Frequency weighted IoU: 0.7448, Loss: 0.5282
Epoch [14/50], Step [10/24], Loss: 0.4613
Epoch [14/50], Step [20/24], Loss: 0.4709
Pixel accuracy: 0.8479, Mean IoU: 0.2489, Frequency weighted IoU: 0.7502, Loss: 0.5259

Epoch [15/50], Step [10/24], Loss: 0.4273
Epoch [15/50], Step [20/24], Loss: 0.4684
Pixel accuracy: 0.8533, Mean IoU: 0.2648, Frequency weighted IoU: 0.7520, Loss: 0.5071
Epoch [16/50], Step [10/24], Loss: 0.4731
Epoch [16/50], Step [20/24], Loss: 0.4556
Pixel accuracy: 0.8510, Mean IoU: 0.2553, Frequency weighted IoU: 0.7560, Loss: 0.5150
Epoch [17/50], Step [10/24], Loss: 0.4418
Epoch [17/50], Step [20/24], Loss: 0.4419
Pixel accuracy: 0.8598, Mean IoU: 0.2647, Frequency weighted IoU: 0.7643, Loss: 0.5007
Epoch [18/50], Step [10/24], Loss: 0.4168
Epoch [18/50], Step [20/24], Loss: 0.3899
Pixel accuracy: 0.8646, Mean IoU: 0.2879, Frequency weighted IoU: 0.7699, Loss: 0.4586
Epoch [19/50], Step [10/24], Loss: 0.3803
Epoch [19/50], Step [20/24], Loss: 0.3695
Pixel accuracy: 0.8677, Mean IoU: 0.3011, Frequency weighted IoU: 0.7790, Loss: 0.4507
Epoch [20/50], Step [10/24], Loss: 0.3698
Epoch [20/50], Step [20/24], Loss: 0.3623
Pixel accuracy: 0.8635, Mean IoU: 0.2969, Frequency weighted IoU: 0.7766, Loss: 0.4625
Epoch [21/50], Step [10/24], Loss: 0.3851
Epoch [21/50], Step [20/24], Loss: 0.3340
Pixel accuracy: 0.8498, Mean IoU: 0.2913, Frequency weighted IoU: 0.7593, Loss: 0.5104
Epoch [22/50], Step [10/24], Loss: 0.3615
Epoch [22/50], Step [20/24], Loss: 0.3854
Pixel accuracy: 0.8714, Mean IoU: 0.3259, Frequency weighted IoU: 0.7846, Loss: 0.4327
Epoch [23/50], Step [10/24], Loss: 0.3502
Epoch [23/50], Step [20/24], Loss: 0.4824
Pixel accuracy: 0.8362, Mean IoU: 0.2614, Frequency weighted IoU: 0.7281, Loss: 0.6016
Epoch [24/50], Step [10/24], Loss: 0.4685
Epoch [24/50], Step [20/24], Loss: 0.4277
Pixel accuracy: 0.8613, Mean IoU: 0.3057, Frequency weighted IoU: 0.7711, Loss: 0.4665
Epoch [25/50], Step [10/24], Loss: 0.4129
Epoch [25/50], Step [20/24], Loss: 0.3695
Pixel accuracy: 0.8648, Mean IoU: 0.3024, Frequency weighted IoU: 0.7785, Loss: 0.4602
Epoch [26/50], Step [10/24], Loss: 0.3495
Epoch [26/50], Step [20/24], Loss: 0.3675
Pixel accuracy: 0.8707, Mean IoU: 0.3180, Frequency weighted IoU: 0.7843, Loss: 0.4331
Epoch [27/50], Step [10/24], Loss: 0.3292
Epoch [27/50], Step [20/24], Loss: 0.3359
Pixel accuracy: 0.8677, Mean IoU: 0.3180, Frequency weighted IoU: 0.7793, Loss: 0.4503
Epoch [28/50], Step [10/24], Loss: 0.3388
Epoch [28/50], Step [20/24], Loss: 0.3086
Pixel accuracy: 0.8774, Mean IoU: 0.3337, Frequency weighted IoU: 0.7936, Loss: 0.4212
Epoch [29/50], Step [10/24], Loss: 0.3096
Epoch [29/50], Step [20/24], Loss: 0.3106

Pixel accuracy: 0.8785, Mean IoU: 0.3418, Frequency weighted IoU: 0.7940, Loss: 0.4151
Epoch [30/50], Step [10/24], Loss: 0.3364
Epoch [30/50], Step [20/24], Loss: 0.3055
Pixel accuracy: 0.8791, Mean IoU: 0.3475, Frequency weighted IoU: 0.7964, Loss: 0.3970
Epoch [31/50], Step [10/24], Loss: 0.3132
Epoch [31/50], Step [20/24], Loss: 0.2899
Pixel accuracy: 0.8791, Mean IoU: 0.3372, Frequency weighted IoU: 0.7969, Loss: 0.4011
Epoch [32/50], Step [10/24], Loss: 0.3009
Epoch [32/50], Step [20/24], Loss: 0.3199
Pixel accuracy: 0.8700, Mean IoU: 0.3192, Frequency weighted IoU: 0.7864, Loss: 0.4377
Epoch [33/50], Step [10/24], Loss: 0.3304
Epoch [33/50], Step [20/24], Loss: 0.3058
Pixel accuracy: 0.8825, Mean IoU: 0.3649, Frequency weighted IoU: 0.8008, Loss: 0.3984
Epoch [34/50], Step [10/24], Loss: 0.3078
Epoch [34/50], Step [20/24], Loss: 0.2810
Pixel accuracy: 0.8812, Mean IoU: 0.3573, Frequency weighted IoU: 0.8000, Loss: 0.4023
Epoch [35/50], Step [10/24], Loss: 0.2788
Epoch [35/50], Step [20/24], Loss: 0.2774
Pixel accuracy: 0.8857, Mean IoU: 0.3758, Frequency weighted IoU: 0.8053, Loss: 0.3884
Epoch [36/50], Step [10/24], Loss: 0.2780
Epoch [36/50], Step [20/24], Loss: 0.2670
Pixel accuracy: 0.8841, Mean IoU: 0.3803, Frequency weighted IoU: 0.8057, Loss: 0.3895
Epoch [37/50], Step [10/24], Loss: 0.2591
Epoch [37/50], Step [20/24], Loss: 0.2752
Pixel accuracy: 0.8829, Mean IoU: 0.3732, Frequency weighted IoU: 0.8058, Loss: 0.3984
Epoch [38/50], Step [10/24], Loss: 0.2747
Epoch [38/50], Step [20/24], Loss: 0.2787
Pixel accuracy: 0.8896, Mean IoU: 0.3929, Frequency weighted IoU: 0.8116, Loss: 0.3774
Epoch [39/50], Step [10/24], Loss: 0.2614
Epoch [39/50], Step [20/24], Loss: 0.2477
Pixel accuracy: 0.8866, Mean IoU: 0.3789, Frequency weighted IoU: 0.8067, Loss: 0.3861
Epoch [40/50], Step [10/24], Loss: 0.2597
Epoch [40/50], Step [20/24], Loss: 0.2507
Pixel accuracy: 0.8893, Mean IoU: 0.4008, Frequency weighted IoU: 0.8099, Loss: 0.3726
Epoch [41/50], Step [10/24], Loss: 0.2387
Epoch [41/50], Step [20/24], Loss: 0.2410
Pixel accuracy: 0.8867, Mean IoU: 0.3865, Frequency weighted IoU: 0.8086, Loss: 0.4002
Epoch [42/50], Step [10/24], Loss: 0.2747
Epoch [42/50], Step [20/24], Loss: 0.3041
Pixel accuracy: 0.8736, Mean IoU: 0.3478, Frequency weighted IoU: 0.7882, Loss: 0.4709
Epoch [43/50], Step [10/24], Loss: 0.2998
Epoch [43/50], Step [20/24], Loss: 0.2871
Pixel accuracy: 0.8841, Mean IoU: 0.3675, Frequency weighted IoU: 0.8038, Loss: 0.4061
Epoch [44/50], Step [10/24], Loss: 0.2529

Epoch [44/50], Step [20/24], Loss: 0.2853
Pixel accuracy: 0.8904, Mean IoU: 0.3976, Frequency weighted IoU: 0.8144, Loss: 0.3684
Epoch [45/50], Step [10/24], Loss: 0.2449
Epoch [45/50], Step [20/24], Loss: 0.2449
Pixel accuracy: 0.8935, Mean IoU: 0.4105, Frequency weighted IoU: 0.8160, Loss: 0.3597
Epoch [46/50], Step [10/24], Loss: 0.2360
Epoch [46/50], Step [20/24], Loss: 0.2322
Pixel accuracy: 0.8913, Mean IoU: 0.4064, Frequency weighted IoU: 0.8116, Loss: 0.3683
Epoch [47/50], Step [10/24], Loss: 0.2227
Epoch [47/50], Step [20/24], Loss: 0.2302
Pixel accuracy: 0.8933, Mean IoU: 0.4222, Frequency weighted IoU: 0.8199, Loss: 0.3616
Epoch [48/50], Step [10/24], Loss: 0.2205
Epoch [48/50], Step [20/24], Loss: 0.2205
Pixel accuracy: 0.8961, Mean IoU: 0.4167, Frequency weighted IoU: 0.8222, Loss: 0.3601
Epoch [49/50], Step [10/24], Loss: 0.2256
Epoch [49/50], Step [20/24], Loss: 0.2170
Pixel accuracy: 0.8978, Mean IoU: 0.4180, Frequency weighted IoU: 0.8222, Loss: 0.3712
Epoch [50/50], Step [10/24], Loss: 0.2059
Epoch [50/50], Step [20/24], Loss: 0.2226
Pixel accuracy: 0.8957, Mean IoU: 0.4114, Frequency weighted IoU: 0.8231, Loss: 0.3528
====================
Finished Training, evaluating the model on the test set
Pixel accuracy: 0.8648, Mean IoU: 0.3513, Frequency weighted IoU: 0.7782, Loss: 0.4698
====================
Visualizing the model on the test set, the results will be saved in the vis/ directory