System Summary

Existing Work

The system is based on the work by Cornia et al. (2018) titled "Predicting human eye fixations via an LSTM-based saliency attentive model." This work can be referenced for more details:

- Marcella Cornia, Lorenzo Baraldi, Giuseppe Serra, and Rita Cucchiara. Predicting human eye fixations via an LSTM-based saliency attentive model. IEEE Transactions on Image Processing, 27(10):5142–5154, October 2018. ISSN 1057-7149. doi: 10.1109/TIP.2018.2851672.

Modifications

The primary modifications compared to the reference work include:

1. The use of the `fcn_resnet101` architecture from PyTorch's torchvision library instead of an LSTM-based model.
2. Implementation of a custom 2D Gaussian kernel for smoothing predictions.
3. Addition of dropout layers to prevent overfitting.

Loss Function and Best Loss Value

The Mean Squared Error (MSE) loss function was used. The best loss value achieved on the validation data was:

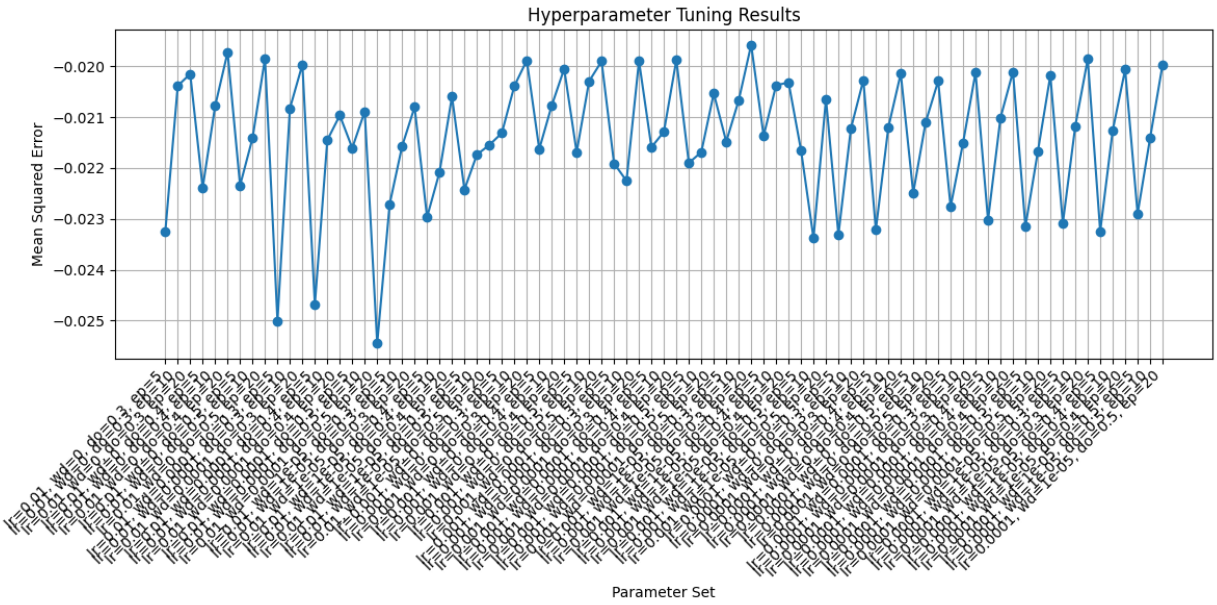
- **Validation RMSE:** 0.057930704206228256 (Fold 5)

Number of Epochs: The network was trained for 5 epochs.

Hyperparameters

The best hyperparameters were found using the grid search. The hyperparameters used in the training are as follows:

- **Learning Rate (lr):** 0.0001
- **Weight Decay:** 1e-4
- **Dropout:** 0.3
- **Batch Size:** 64
- **Optimizer:** Adam



Data Augmentation and Training Procedures

- **Transformations:**
  - For images: `ToTensor()`
  - For fixations: `Grayscale(), ToTensor()`
- **Training Procedure:**
  1. Data loading and preparation using `DataLoader`.
  2. Model training with the Adam optimizer and MSE loss function.
  3. Cross-validation using K-Fold with 5 splits.
  4. Model evaluation based on RMSE, accuracy, and ROC-AUC metrics.

Evaluation Metrics

The average scores obtained across the folds are as follows:

- **Average Validation RMSE:** 0.06320345177207625
- **Average Accuracy:** 0.7668085962266972
- **Average ROC-AUC:** 0.8247200918570029

Predicted images