RWU Hochschule Ravensburg-Weingarten University of Applied Sciences

Faculty for Electrical Engineering and Computer Science

Task Report

Image Fusion for Enhanced Shape Classification

Guided by:

Mr. Felix Berens

*S*ubmitted by:

Akshay Chobe-36316

Master-Mechatronics

27 June 2024

Introduction

In this report, a low and high level fusion strategy is used to improve the classification performance of a synthetic dataset. The 1000 instances in the dataset each have three 32x32 images that represent one of the four geometric shapes (pentagon, circle, square, or triangle). Three distinct background types—gradient, noise, and spotlight—are used to create the images. To classify these shapes, a convolutional neural network called LeNet-5 is implemented. To compare performance improvements, 3 LeNet5 models are trained and evaluated on the images without fusion at first and then another LeNet5 model is trained and evaluated on fused image dataset obtained from the defined fusion strategy. Finally the increased accuracy of the CNN model is achieved for the shape classification task. At the end confusion matrices are plot to demonstrate these accuracies.

Step 1: Change LeNet5 Architecture

Yann LeCun created LeNet-5 in 1998 with the intention of using the MNIST dataset to recognize handwritten digits[1]. The layers that make up the original architecture and the modified LeNet architecture suitable for the Image Fusion dataset are differentiated through figures below along with the description of modified layers:

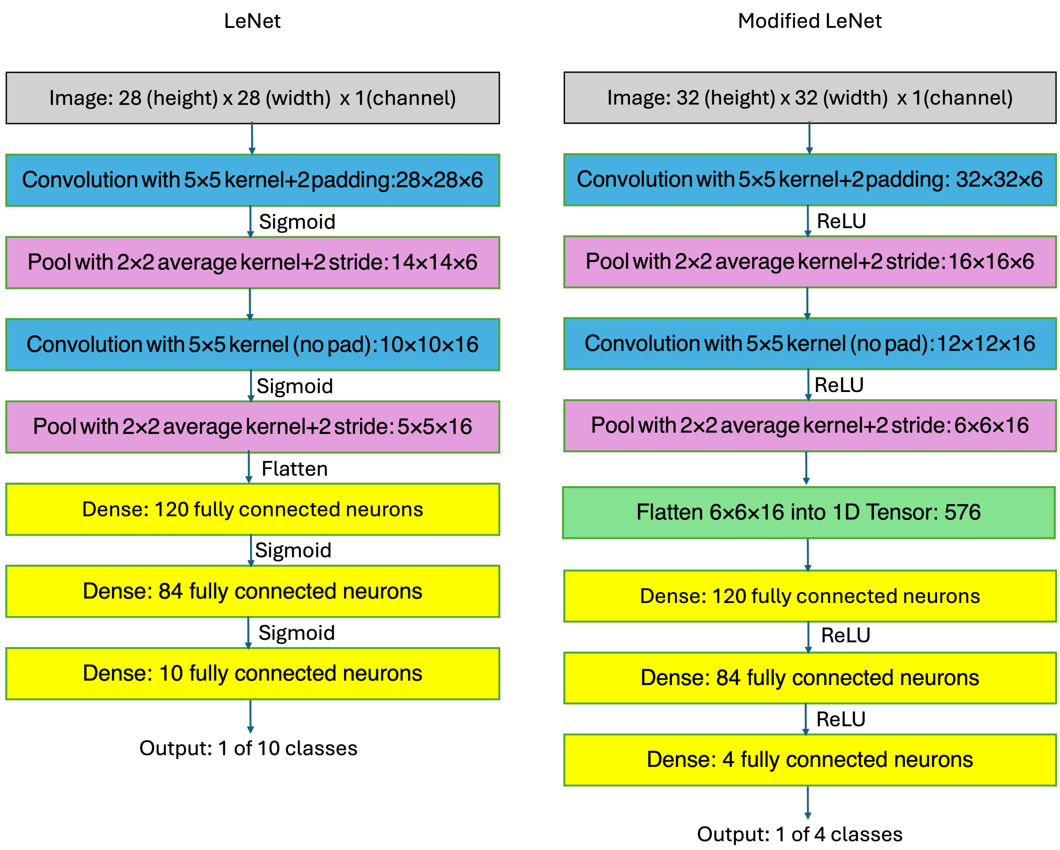

Figure 1: Original LeNet5 model (left) and Modified LeNet5 model (right)

1. **Input Layer:** Adapted to accept 32x32 Grayscale images (1 input channel).
2. **First Convolutional Layer:** Kept 1 input channel as it is because the original images are grayscale. Applied six 5x5 filters to produce six 32x32 feature maps. ReLU activation was chosen to introduce non-linearity and improve training convergence over sigmoid.
3. **First Subsampling Layer:** Average pooling to reduce feature maps to 16x16.
4. **Second Convolutional Layer:** Sixteen 5x5 filters to produce sixteen 12x12 feature maps.
5. **Second Subsampling Layer:** Average pooling to reduce feature maps to 6x6.
6. **Flatten Layer:** Output flattened to 6x6x16 = 576 neurons for input to fully connected layers.
7. **Fully Connected Layers:** Two fully connected layers with 120 and 84 neurons were implemented with ReLU activation function chosen for its ability to handle complex relationships in data and accelerate convergence during training.
8. **Output Layer:** 4 neurons corresponding to the 4 classes (geometric shapes), using softmax activation for multi-class classification to output probabilities.

   These changes enable LeNet-5 to handle larger, RGB images and distinguish between the four classes effectively.

Step 2: Train and Evaluate LeNet without Fusion

- **Dataset Preparation:** The synthetic dataset consists of 3 folders each having 1000 images of size 32x32, with different background but the same geometric shape. For every background type—gradient, noise, and spotlight; the dataset was divided into training (80%) and testing (20%) sets.
- **Training:** For each background type, LeNet-5 was trained separately on each type of image directory img1- Model 1, img2- Model 2 and img3- Model 3. Cross-entropy loss was chosen for its effectiveness in multi-class classification tasks like identifying geometric shapes, while Adam optimizer with a learning rate of 0.001 was selected for efficient convergence and adaptation to varying gradients during training, optimizing the model's performance on this specific dataset. A smaller batch size of 40 and epochs used are 50 to enhance generalization and to prevent overfitting.
- **Evaluation:** To establish a baseline for comparison with the fusion strategies, the model's accuracy on the testing set for each background type was noted with confusion matrices and visualised below:

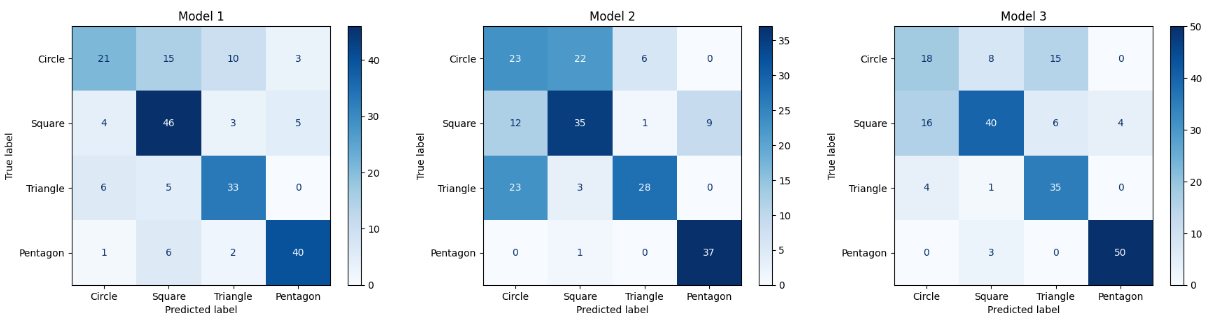

Figure 2: Confusion matrices for individual models without fusion

Step 3: Implementing fusion Strategy

To integrate data from the three different image types, two main fusion techniques were taken into consideration:

1. Low-Level Fusion: Directly combining the three images' pixel values[2].

- **Pixel-based Average Fusion:** The average of the matching pixel values from each of the three original images is used to determine the value of each pixel in the fused image.
- **Pixel-based Maximum Fusion:** In the fused image, every pixel value represents the highest value of its corresponding pixels in the three original images.
- **Pixel-based Minimum Fusion:** In the three original images, each pixel value is the minimum of its corresponding pixel values in the fused image.

After all, the average fusion method produced a well-balanced combination of the three images, so it was chosen for additional testing. A new directory was created to store the fused images dataset, which is then used for training new LeNet model after low level fusion. Figures below show two randomly selected images for visualisation of low level image fusion:

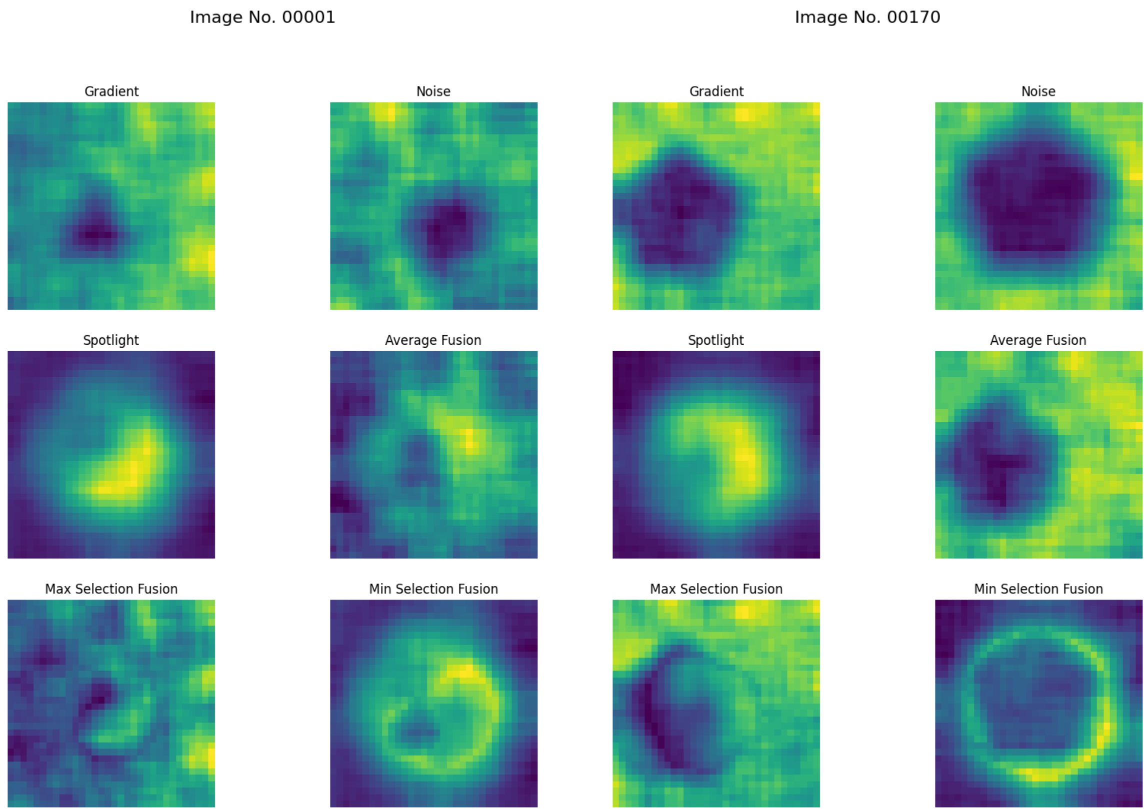

Figure 3: Low level fusion Implementation

2. High-Level Fusion: Merging the results of different models that have been trained on every kind of image.

- **Majority Voting:** Among all four models, the final classification is decided by majority vote. The prediction of the fused model is used as a tiebreaker in the event of a tie.

Step 4: Train and Evaluate LeNet with Fusion

Low-Level Fusion Training and Evaluation: A new LeNet-5 model was trained on the fused dataset, which was produced by the average fusion method. The training configuration was kept the same, using the Adam optimizer, cross-entropy loss, and 50 epochs to maintain the transparency in comparison. Subsequently, the combined model was assessed on the testing dataset, and its precision was compared with the individual models trained on each kind of background without fusion.

High-Level Fusion Training and Evaluation: Using majority voting, the predictions from the three distinct models which trained on a different type of image sets were combined with that of the low-level fusion model. This resulted in comparing the mostly predicted class with the true label, giving high classification probabality with the used fusion strategy.

Results

The accuracy of the models trained with the fusion strategy was significantly higher than those trained on individual image types, demonstrating the effectiveness of the fusion strategy. The deep learning model's performance evaluation tools like accuracy and confusion matrix are implemented. The accuracy of pixel based average selection fusion is also calculated for the evaluation of the significance of integrating low and high level fusion strategies.

Models	Accuracy
Model 1 - Gradient Background	70.0%
Model 2 - Noise Background	61.5%
Model 3 - Spotlight Background	71.5%
Low-Level Fusion (Average Selection)	77.5%
Low and High-Level Fusion (Average+Majority Voting)	82.5%

Table 1: Model accuracy results for different background types and fusion methods.

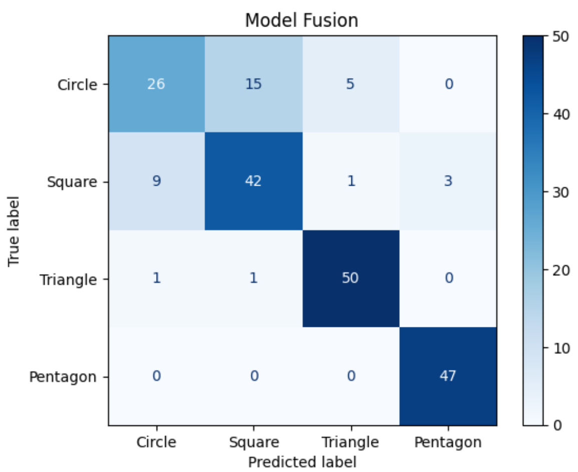

Figure 4: Confusion Matrix for Fusion Model

Conclusion

The fusion strategy, both at the low-level and high-level, significantly improved the classification accuracy of the LeNet-5 model. The low-level fusion method (average fusion) provided a straightforward yet effective approach to integrating information from the three image types. High-level fusion further enhanced accuracy through majority voting, leveraging the strengths of models trained on different image backgrounds.

Future work could investigate additional fusion techniques, like mid-level fusion where the features of the images will be evaluated rather than only pixel values. Hyperparameters can be tuned to further enhance the classification task, testing on a larger and more varied dataset might provide more details about the robustness and scalability of the fusion strategies, for this data augmentation techniques can be used. Additionally more complex architecture like VGG, AlexNet or GoogleNet can be used for the image classsification task which produce overall highest accuracies for classification[3].

References

1. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. (1998). "LeNet-5." *Wikipedia, The Free Encyclopedia*. Wikimedia Foundation. Available at: https://en.wikipedia.org/wiki/LeNet

2. Kalamkar, S., Geetha. M., "Multimodal image fusion: A systematic review." *Decision Analytics Journal*, Volume 9, 2023, Article 100327. Elsevier. Available at: www.elsevier.com/locate/dajour

3. Xi, P.; Shu, C.; Goubran, R. "Abnormality detection in mammography using deep convolutional neural networks." *IEEE International Symposium on Medical Measurements and Applications (MeMeA)*., Rome, Italy. Available at: https://shorturl.at/1D5S5

4. Schutera, M., Hafner, F. "Why does my Neural Network not learn?: Coffee Table Solutions for Deep Learning", Paperback. *English Edition*