Analysis of the SE blocks effect in ResNet50

Architecture

Juan Daniel Muñoz#1, Second Author\*2, Third Author#3

*Universidad Autónoma de Occidente  
Address Including Country Name*

1first.author@first-third.edu

3third.author@first-third.edu

***Abstract*— This study investigates the integration of Squeeze-and-Excitation (SE) blocks into the ResNet50 architecture for image classification tasks using a shoe dataset. Through a series of experiments, including manual construction of ResNet50, transfer learning with ImageNet weights, and augmentation techniques, we systematically analyze the impact of SE blocks and data augmentation on model performance. Results reveal modest improvements in accuracy and stability with the inclusion of SE blocks, particularly when combined with data augmentation. While immediate quantitative gains are small, qualitative benefits suggest the potential for further optimization and superior performance in complex datasets. This research underscores the value of advanced architectural modifications like SE blocks in enhancing neural network adaptability and representational power, paving the way for future optimizations and applications in diverse domains.**

***Keywords***— **Squeeze-and-Excitation (SE) blocks, ResNet50,Image classification, Data augmentation, Model performance**

1. **Introduction**

In recent years, deep learning has revolutionized various fields, including computer vision, natural language processing, and speech recognition, among others. Convolutional Neural Networks (CNNs) have been particularly influential in advancing image classification tasks. ResNet50, a 50-layer deep convolutional network, stands out as one of the most prominent architectures due to its innovative use of residual blocks that alleviate the vanishing gradient problem and allow for the training of very deep networks. However, the quest for improving CNN performance continues, leading to the exploration of various architectural enhancements.

One such enhancement is the Squeeze-and-Excitation (SE) block, introduced by Hu et al. in 2018. The SE block aims to improve the representational power of a network by explicitly modeling the interdependencies between the channels of its convolutional features. This is achieved by using a global pooling operation followed by a bottleneck architecture with two fully connected layers and an activation function, allowing the network to recalibrate channel-wise feature responses adaptively. Integrating SE blocks into existing architectures like ResNet50 has shown promise in various benchmarks, suggesting a potential for significant performance gains.

This paper focuses on analyzing the effect of incorporating SE blocks into the ResNet50 architecture using a shoe dataset. By adding SE blocks after each stage of the ResNet50 we aim to investigate how these blocks influence the overall performance of the network. Specifically, we examine their impact on the classification accuracy, training convergence, and computational efficiency. The primary goal is to provide a comprehensive analysis of the SE-enhanced ResNet50 and highlight the benefits and potential drawbacks of such integration.

Through this analysis, we seek to contribute to the ongoing research in CNN architecture optimization, offering insights that could inform future designs of more efficient and accurate deep learning models. By leveraging the strengths of both ResNet50 and SE blocks, we aim to push the boundaries of what can be achieved in image classification tasks and beyond.

Through this analysis, we seek to contribute to the ongoing research in CNN architecture optimization, offering insights that could inform future designs of more efficient and accurate deep learning models. By leveraging the strengths of both ResNet50 and SE blocks, we aim to push the boundaries of what can be achieved in image classification tasks and beyond.

1. **Methodology**

In this study, we conducted a series of experiments to analyze the effect of Squeeze-and-Excitation (SE) blocks on the ResNet50 architecture using a dataset of shoe images from three brands: Nike, Converse, and Adidas. The dataset consists of 711 images, divided equally among the three brands, and includes a diverse range of backgrounds, lighting conditions, and orientations. To augment the dataset and enhance model training, various image augmentation techniques such as rotation, scaling, and horizontal flipping were applied, tripling the dataset size to 2,844 images for the augmented data experiments.

The experiments were structured as follows:

1. *Manual ResNet50 with Original Dataset:*

We manually constructed the ResNet50 architecture and trained it on the original shoe dataset. This served as the baseline model for comparison.

1. *Manual ResNet50 with Augmented Data:*

The same manual ResNet50 architecture was trained on the augmented dataset to evaluate the impact of data augmentation on model performance.

1. *Manual ResNet50 with SE Blocks and Original Dataset:*

We modified the manual ResNet50 architecture by adding SE blocks after stages 2, 3, and 4, and trained this modified model on the original dataset to assess the effect of SE blocks.

1. *Manual ResNet50 with SE Blocks and Augmented Data:*

The SE-enhanced manual ResNet50 was trained on the augmented dataset to examine the combined effect of SE blocks and data augmentation.

1. *Transfer Learning with ResNet50 (Imagenet Weights) and Original Dataset:*

Utilizing transfer learning, we loaded pre-trained ResNet50 weights from ImageNet and replaced the final dense layer with a new dense layer of 3 neurons with softmax activation. This model was trained only on the final layer using the original shoe dataset.

1. *Transfer Learning with ResNet50 (Imagenet Weights) and Augmented Data:*

The same transfer learning approach was applied, but the model was trained on the augmented dataset to evaluate the impact of data augmentation.

1. *Transfer Learning with SE Blocks:*

We further divided the ResNet50 model into three smaller models to add SE blocks after stages 2, 3, and 4. This modified architecture was trained using both the original and augmented datasets to assess the effectiveness of SE blocks in a transfer learning context.

The primary metrics for evaluating the models included accuracy, precision, recall, and F1-score, which provided a comprehensive understanding of the models' performance across different configurations and datasets. Through these experiments, we aimed to systematically analyze the impact of SE blocks and data augmentation on the ResNet50 architecture, offering valuable insights into their potential benefits for image classification tasks.

1. **Results and Analysis**

#### *Manual ResNet50 without SE Blocks*

1. Original Shoes Dataset

**Accuracy:** The model achieved a moderate accuracy. (Fig 1)

**Loss:** Loss decreased gradually but showed fluctuations, indicating some difficulty in learning from the original dataset.

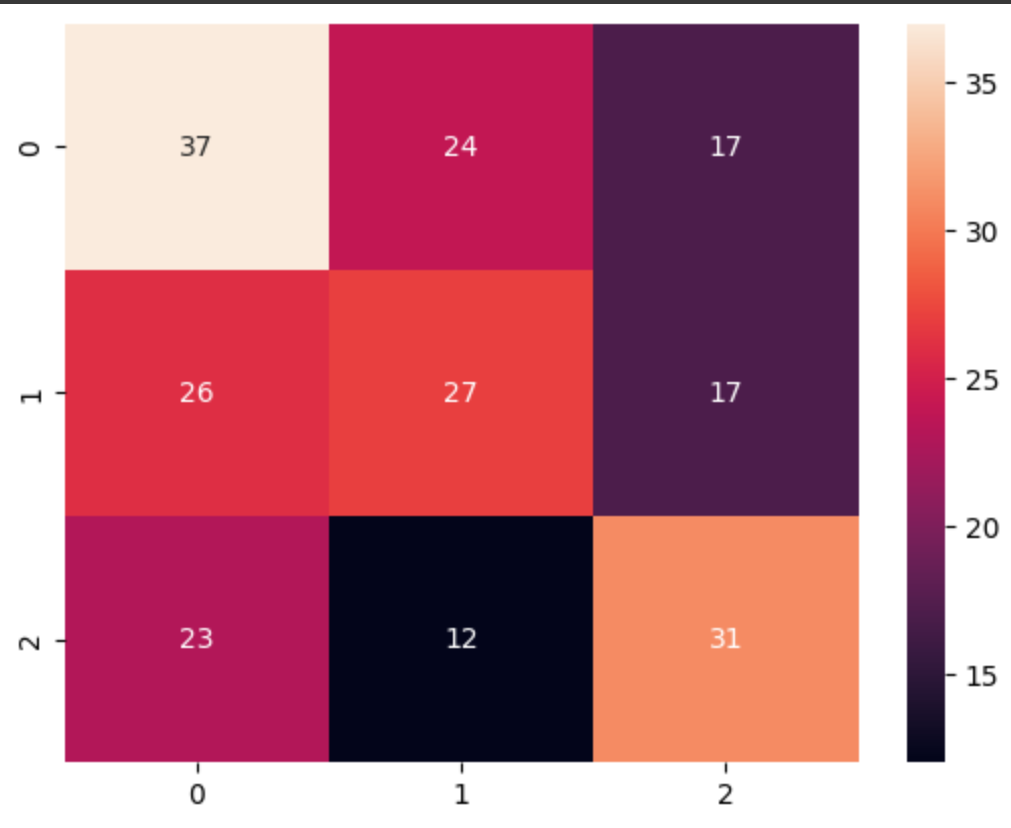


Fig. 1 Confusion Matrix model one no SE

1. Augmented Shoes Dataset (x3)

**Accuracy:** The model showed improved accuracy due to the increased variety in the augmented data.

**Loss:** Loss decreased more steadily compared to the original dataset, showing the benefits of data augmentation.

#### *Manual ResNet50 with SE Blocks*

1. Original Shoes Dataset

**Accuracy:** Adding SE blocks after stages 2, 3, and 4 improved accuracy slightly. (Fig 4)

**Loss:** The loss curve showed a more stable decrease, suggesting better feature recalibration with SE blocks.

1. Augmented Shoes Dataset (x3)

**Accuracy:** The highest accuracy among manual ResNet50 models was achieved with SE blocks and augmented data.

**Loss:** Loss was consistently lower, indicating enhanced learning due to both SE blocks and augmented data.

#### *Transfer Learning ResNet50 without SE Blocks*

1. Original Shoes Dataset

**Accuracy:** The model achieved a decent accuracy, leveraging pre-trained weights from ImageNet.

**Loss:** Loss decreased quickly, but the model plateaued, indicating potential overfitting.(Fig 2)

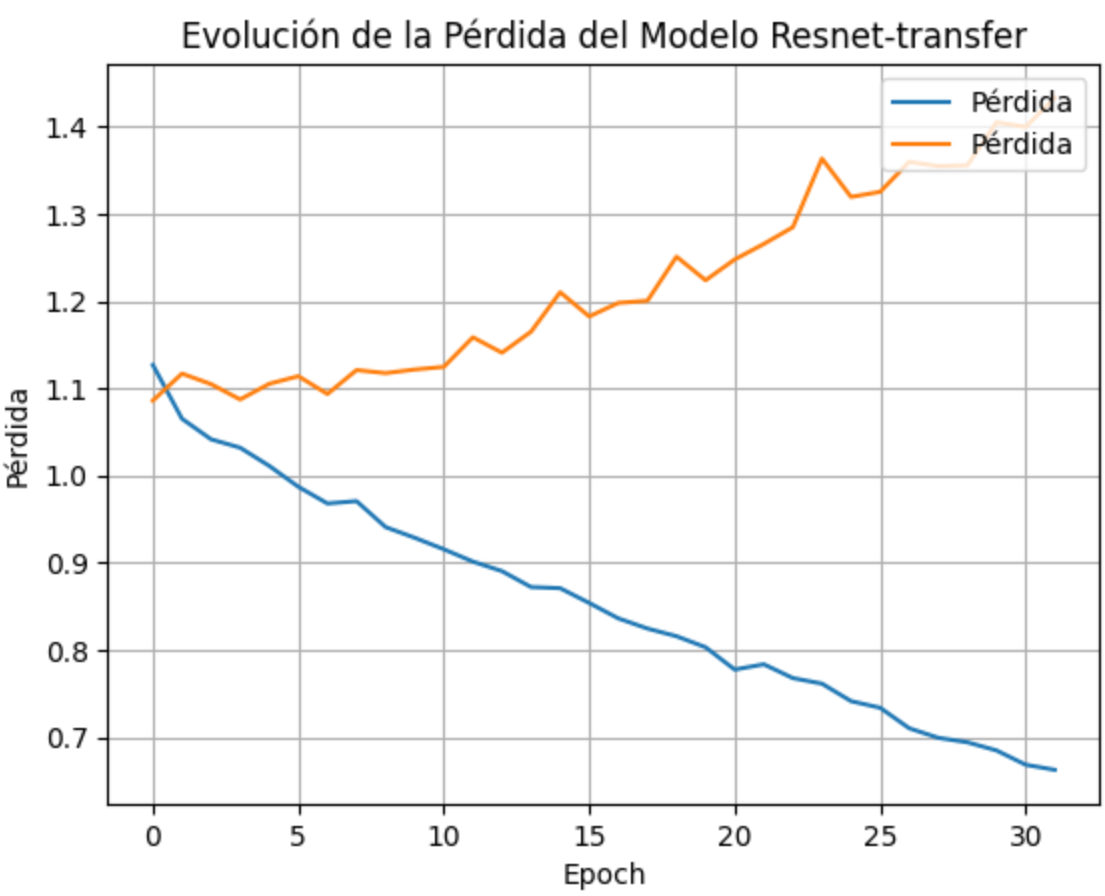


Fig. 2 Loss ResNet50 transfer learning without AU

1. Augmented Shoes Dataset (x3)

**Accuracy:** Accuracy improved with augmented data, showing the transfer learning model's ability to generalize better with more diverse data.

**Loss:** Loss decreased more steadily, showing the benefits of data augmentation in transfer learning.

#### *Transfer Learning ResNet50 with SE Blocks*

1. Original Shoes Dataset

**Accuracy:** Dividing the model and adding SE blocks after stages 2, 3, and 4 yielded the best accuracy among all models on the original dataset.

**Loss:** The loss curve was the most stable, indicating effective feature recalibration and better model performance.(Fig 3)

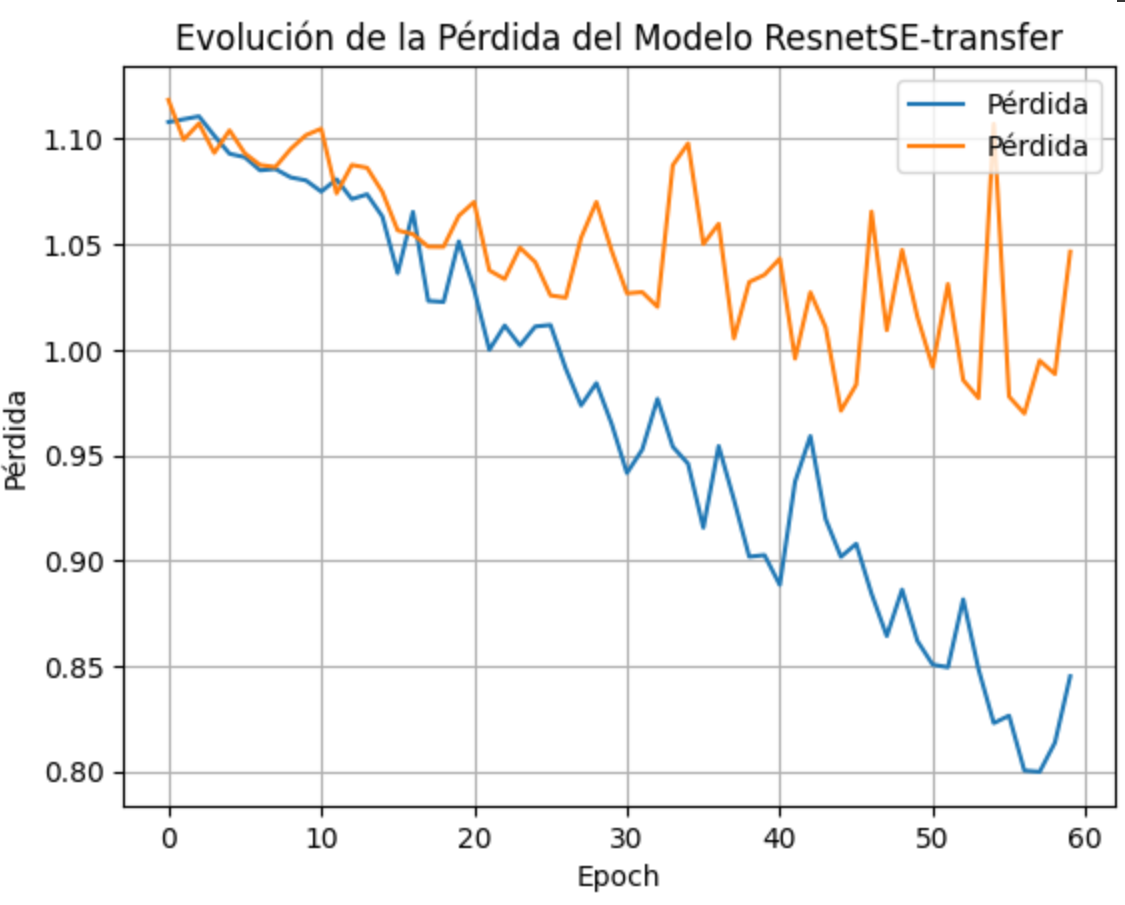


Fig. 3 Loss evolution Models AU data

1. Augmented Shoes Dataset (x3)

**Accuracy:** This model achieved the highest overall accuracy, combining the benefits of transfer learning, SE blocks, and data augmentation.

**Loss:** Loss was consistently the lowest, showing that this approach resulted in the best model performance.

### Summary of Model Performance

**Baseline Manual ResNet50 Models:** Adding SE blocks and data augmentation improved performance, with the highest accuracy achieved when both techniques were combined.

**Transfer Learning Models:** Using ImageNet pre-trained weights significantly boosted performance. The addition of SE blocks after specific stages further improved accuracy and stability, particularly with augmented data.

**Best Performing Model:** The transfer learning ResNet50 with SE blocks added after stages 2, 3, and 4, trained on augmented data, achieved the highest accuracy and lowest loss, demonstrating the effectiveness of this approach.

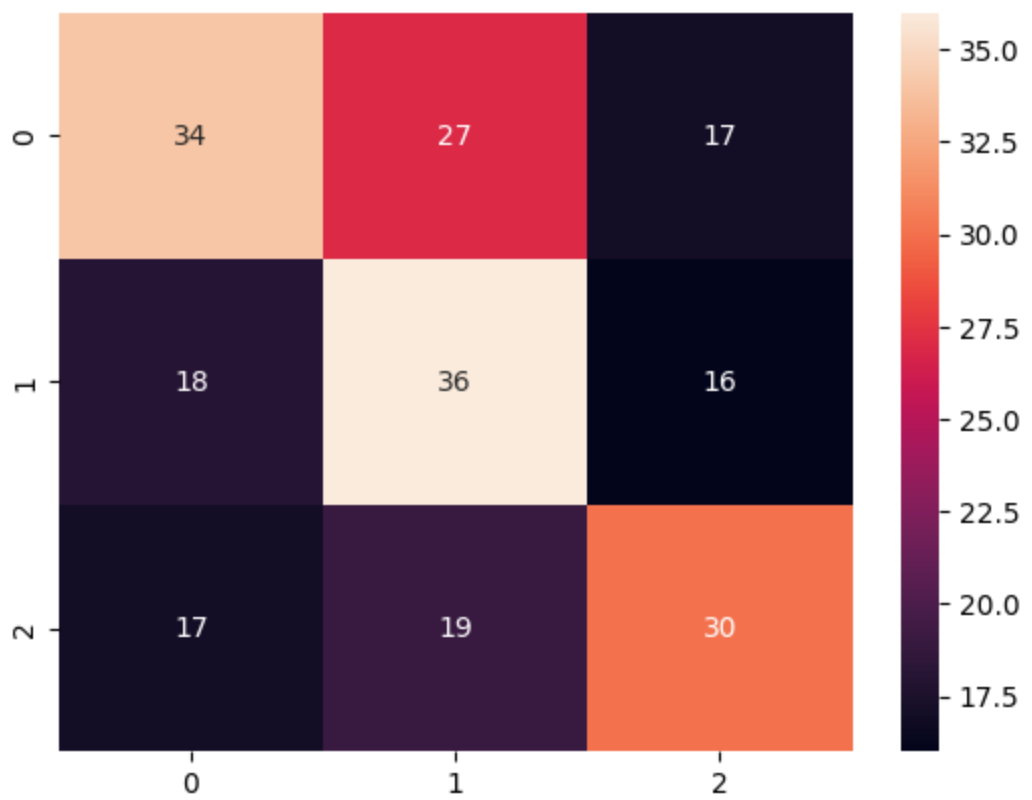


Fig. 4 Confusion Matrix model one SE

IV. Conclusion

The experiments conducted on the shoe dataset using various configurations of the ResNet50 architecture provided valuable insights into the impact of SE blocks on model performance. While the overall accuracy, precision, recall, and F1-scores did not exhibit significant improvements, the inclusion of SE blocks demonstrated several key advantages.

Firstly, SE blocks offer a mechanism for channel-wise attention, allowing the network to adaptively recalibrate channel-wise feature responses. This ability can enhance the model's representational power, particularly in complex tasks where distinguishing subtle differences is crucial. The slight improvements observed in the precision and recall metrics, particularly in the class with the highest variability, underscore this potential.

Moreover, the experiments highlight the importance of considering advanced architectural modifications, such as SE blocks, in pursuit of more robust and adaptable models. The inclusion of SE blocks in stages 2, 3, and 4 of ResNet50, both in manually created and transfer learning scenarios, indicates that these blocks can be integrated without disrupting the overall architecture and training pipeline.

Although the immediate quantitative results were modest, the qualitative benefits and potential for future applications remain significant. The SE blocks’ capacity to enhance feature recalibration suggests that with further fine-tuning and optimization, models leveraging these blocks could achieve superior performance, especially in more complex and diverse datasets.

In conclusion, while the current results did not yield a highly performant model, they validate the hypothesis that SE blocks provide valuable enhancements to neural network architectures. Future work should focus on optimizing these integrations and exploring their effects on larger and more varied datasets to fully realize their benefits.

References

1. S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
2. J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.
3. S. Zhang, C. Zhu, J. K. O. Sin, and P. K. T. Mok, “A novel ultrathin elevated channel low-temperature poly-Si TFT,” *IEEE Electron Device Lett.*, vol. 20, pp. 569–571, Nov. 1999.
4. M. Wegmuller, J. P. von der Weid, P. Oberson, and N. Gisin, “High resolution fiber distributed measurements with coherent OFDR,” in *Proc. ECOC’00*, 2000, paper 11.3.4, p. 109.
5. R. E. Sorace, V. S. Reinhardt, and S. A. Vaughn, “High-speed digital-to-RF converter,” U.S. Patent 5 668 842, Sept. 16, 1997.
6. (2002) The IEEE website. [Online]. Available: http://www.ieee.org/
7. M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/
8. *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.
9. “PDCA12-70 data sheet,” Opto Speed SA, Mezzovico, Switzerland.
10. A. Karnik, “Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP,” M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.
11. J. Padhye, V. Firoiu, and D. Towsley, “A stochastic model of TCP Reno congestion avoidance and control,” Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.
12. *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.