Intelligent information systems

UDC 004.93 **doi:** https://doi.org/10.20998/2522-9052.2023.3.07

Artem Melnychenko, Kostyantyn Zdor

National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine

EFFICIENCY OF SUPPLEMENTARY OUTPUTS IN SIAMESE NEURAL NETWORKS

Abstract. In the world of image analysis, effectively handling large image datasets is a complex challenge that requires using deep neural networks. Siamese neural networks, known for their twin-like structure, offer an effective solution to image comparison tasks, especially when data volume is limited. This research explores the possibility of enhancing these models by adding supplementary outputs that improve classification and help find specific data features. The article shows the results of two experiments using the Fashion MNIST and PlantVillage datasets, incorporating additional classification, regression, and combined output strategies with various weight loss configurations. The results from the experiments show that for simpler datasets, the introduction of supplementary outputs leads to a decrease in model accuracy. Conversely, for more complex datasets, optimal accuracy was achieved through the simultaneous integration of regression and classification supplementary outputs. It should be noted that the observed increase in accuracy is relatively marginal and does not guarantee a substantial impact on the overall accuracy of the model.

Keywords: computer vision; neural networks; Siamese neural networks; image recognition.

Introduction

The realm of image processing is confronted with the challenge of managing high-dimensional data, necessitating the construction of neural networks characterized by profound architectures for optimal efficiency.

Yet, the efficacy of deep neural networks depends on access to huge datasets [1]. Addressing this conundrum, Siamese neural networks emerge as a partial remedy, offering a solution to the data scarcity issue inherent in the deployment of intricate neural architectures.

Siamese neural networks constructed as paired twins within a shared architecture, these networks excel in capturing intricate data representations and discerning nuanced dissimilarities, rendering them particularly adept at tasks involving modest dataset sizes [2].

The architecture of Siamese neural networks consists of a part with the encoding of input data and an algorithm for their comparison [3].

After encoding the image into a multidimensional model, it is possible to calculate the distance between different encoded images [4].

The most common way of calculating the distance between encoded data by the Siamese neural network is the Euclidean distance [5], which is calculated by the formula:

$$D_W\left(\vec{X}_1, \vec{X}_2\right) = \| G_W\left(\vec{X}_1\right) - G_W\left(\vec{X}_2\right) \|_2 ,$$

where $\vec{X}_i$ are input images, G_W is a transformation function, in our case it is a neural network, D_W is the distance between images.

This format of results requires a special function for determining the error – contrastive loss. Contrastive loss calculates the error in the received distance relative to the expected one.

The error is calculated using the formula:

$$L\left(W, Y, \vec{X}_1, \vec{X}_2\right) = \\ = \frac{1-Y}{2}\left(D_W\right)^2 + \frac{Y}{2}\left(max\left(0, m - D_W\right)\right)^2 ,$$

where W are system parameters, Y is the expected distance between images, m is the expected distance between different images.

The architectural model simplifies image clustering even with limited data, crucial for specialized domains; however, accuracy drops can occur due to insufficient domain knowledge and data. Without extra layers, the neural network struggles to select vital domain-specific features, but adding these layers can bolster hyperspace robustness, emphasizing unique attributes. Fine-tuning involves selecting a pre-trained model (preferably domain-specific), freezing low-level abstraction layers, and adding new layers suitable for the task before training [7]; while this enhances outcomes, it doesn't fully address feature and domain context placement issues.

Problem statement. The primary objective of this research paper is to improve the precision and performance of models by strategically guiding their attention toward supplementary features present within comparison images. The existing problem stems from the limitations of Siamese neural networks in effectively utilizing supplementary information during training. To address this issue, the research aims to design an algorithm that enhances the training efficacy of Siamese neural networks by introducing supplementary branches. These supplementary branches will allow the model to handle classification and regression challenges more effectively. The research proposes the creation of variant Siamese neural networks equipped with extra outputs that specifically address both classification and regression tasks. By incorporating these supplementary branches and outputs, the research strives to achieve a more comprehensive and accurate model capable of leveraging a wider range of information for improved performance across various tasks.

Proposed approach

To enhance the training outcomes of Siamese neural networks, we suggest integrating supplementary outputs for classification or attribute-specific searches. Drawing inspiration from the GoogLeNet model, which employed supplementary outputs for classification to address the challenge of inadequate error propagation in expansive models [8].

Consequently, we aimed to evaluate the efficacy of incorporating these supplementary outputs for attribute computation in Siamese neural networks.

To produce these results, specific data alterations are required to capture more overarching traits. Consequently, the added outputs also undertake the role of contrasting broader features or classification tasks. This method demands extensive time for data investigation, feature identification, and crafting intricate datasets to train the model on multiple tasks simultaneously. Through this training, the model is compelled to identify essential features for additional issues at a more basic level, enhancing its accuracy and generalization capacities.

Concurrently, it's vital to monitor the weight distribution across the model's outputs to ensure supplementary branches don't disrupt the primary training.

Experiment 1. Training Siamese neural network with supplementary classification output, using varying loss weights on fashion MNIST dataset

For our initial experiment, we decided to utilize a traditional dataset, comparing the foundational architecture against its modified version enhanced with extra outputs. We chose the Fashion MNIST dataset [9] because it is a widely used benchmark in the field of computer vision, serving as a modern alternative to the traditional handwritten digit recognition dataset (MNIST).

Comprising 70,000 grayscale images spanning 10 different clothing categories, such as T-shirts, trousers, and dresses, the dataset offers a diverse array of fashion items for classification tasks as shown in Fig. 1.

Fig. 1. Example of Fashion MNIST dataset samples

Algorithm 1. Compare training results for a Siamese neural network with supplementary classification output.

Input:
- Fashion Mnist dataset.

Output:
- Trained Siamese neural network
- Trained Siamese neural networks with supplementary classification output, using varying loss weights.

Procedure:

Step 1. Load the dataset and generate pairs for the Siamese neural network.

Step 2. Calculate supplementary outputs.

Step 3. Split dataset to training, validation, and test samples.

Step 4. Transform and augment training samples.

Step 5. Train Siamese neural network

Step 6. Train Siamese neural network with supplementary classification output, using varying loss weights.

For the experiments, two rudimentary Siamese neural networks were developed. The initial model followed a traditional architecture, calculating the distance between encoded images.

Conversely, the second model incorporated an added output featuring a classifier at its termination as shown in Fig. 2. Both architectures utilized identical layers, a contrastive loss computation function used for the base output, and RMSProp was selected as the optimizer [10]. Based on experimental evaluations, 100 epochs were considered sufficient for the model to achieve its optimal performance.

Initial comparisons indicated that the standalone Siamese neural network outperformed in terms of faster convergence and achieving an accuracy rate of 92.3%.

Consequently, adjustments were made by incorporating weights into the model's error computation functions. Subsequent tests, using weight multipliers of 0.5, 0.2, and 0.1 for the supplementary output (Table 1), revealed that reduced error function weights produced superior outcomes [11].

Table 1 – **Comparison of training results with auxiliary classification task**

Additional outputs	Accuracy
No additional outputs	92.3%
Classification, loss weights 100%	90,08%
Classification, loss weights 50%	90,93%
Classification, loss weights 20%	91,75%
Classification, oss weights 10%	92,07%

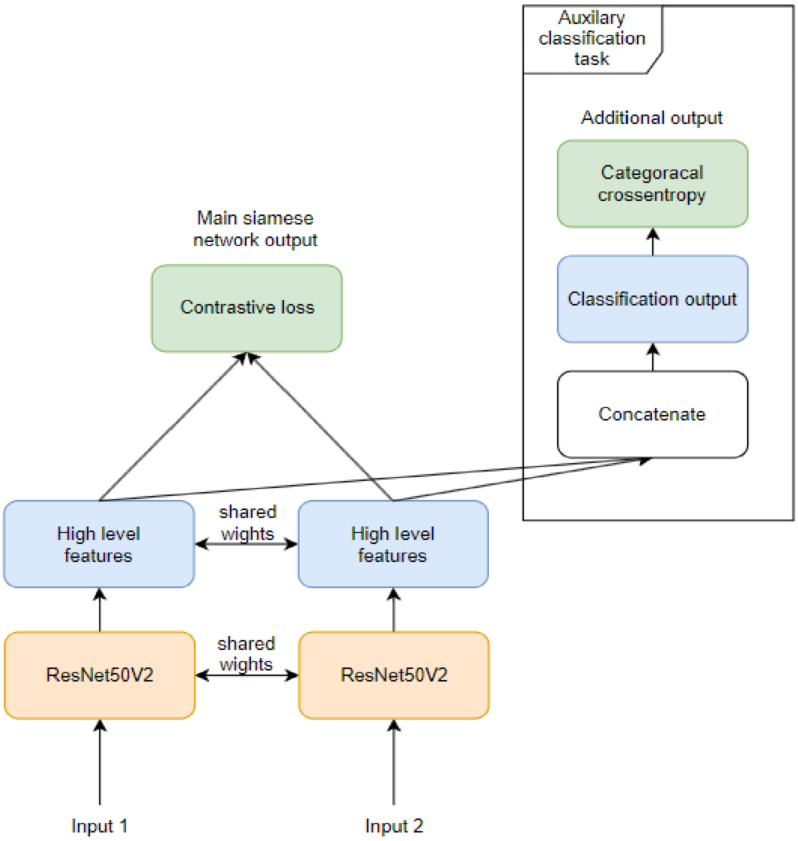

Fig. 2. Siamese neural network architecture with supplementary classification branch

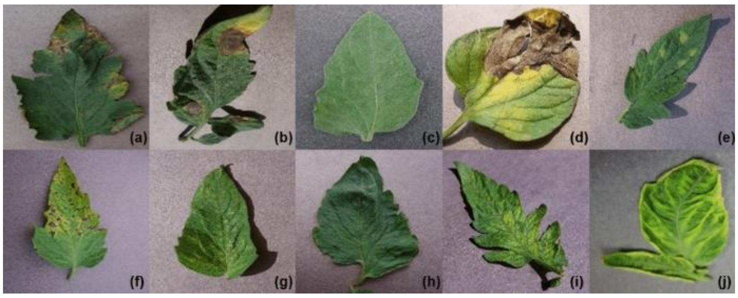

Fig. 3. Example of PlantVillage dataset samples

Experiment 2. Training Siamese neural network with supplementary outputs, using varying loss weights on plantvillage dataset

For the subsequent experiment, we decided to use a more intricate and challenging dataset for our forthcoming investigation. This decision stems from the imperative to enhance the depth and robustness of our study's outcomes, so we utilized a subset of the PlantVillage dataset [12]. This dataset comprises approximately 20,000 images spread across 15 recognition classes as shown in Fig. 3. To identify extra features, the dataset was segmented into hypothetical groups. The first group categorizes the plant type, resulting in three distinct classes: "Pepper," "Potato," and "Tomato." Another classification was based on the health status of the plants – either healthy or diseased. Consequently, we were able to generate supplementary outputs for classifying plant types and a regression output indicating the plant's health status.

The first algorithm was modified and additional steps were added to enhance its capabilities. Specifically, in Step 2, supplementary outputs were calculated for the classification task based on the leaf type, and in Step 3, supplementary outputs were calculated for the regression task based on the health status of the leaf. These additional steps were aimed at providing more comprehensive information and guidance to the Siamese neural network during training, enabling it to better distinguish specific properties, e.g. leaf types and leaf health.

The ResNet50V2 model was selected due to its prior training on the extensive Imagenet dataset, encompassing millions of categorized images. This pretraining bestows the model with robust generalization capabilities and a broad spectrum of visual feature extraction. Additionally, the model stands out for its rapid execution and minimal resource requirements [13].

Subsequent layers were incorporated to interpret features derived from ResNet50V2, forming a vector for a Siamese neural network. These vectors were then directed to an Euclidean distance computation function, with contrastive loss employed for error determination. The model's training converged in a 96.54% accuracy rate.

Progressing further, we applied supplementary outputs, with the first supplementary output designed for regression aiming to guide the model's focus toward leaf conditions. A subsequent model iteration incorporated a classifier as an extra output, targeting the model's attention to letter categorizations.

In the final iteration, a dual-output model was crafted, amalgamating both regression and classifier recognition functionalities from the prior experiments as shown in Fig. 4.

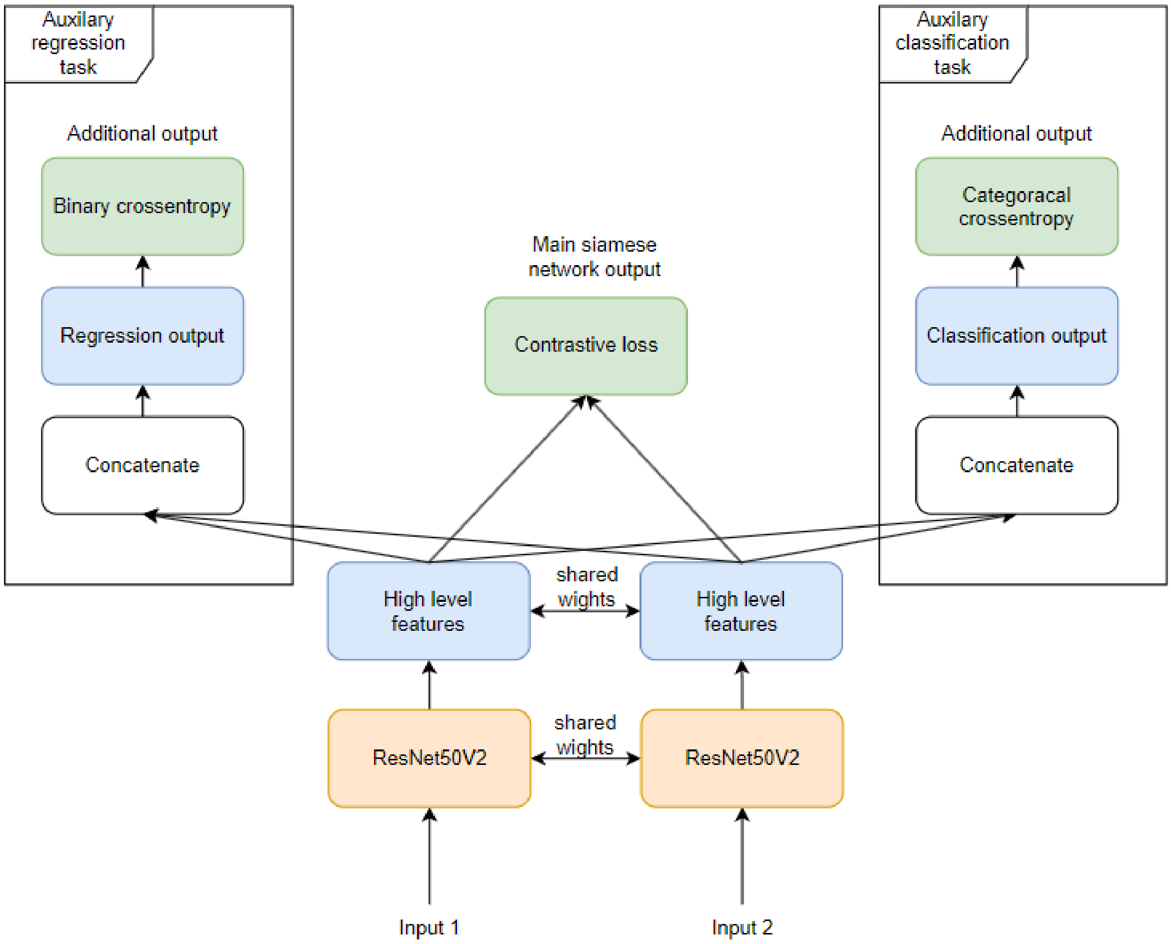

Fig. 4. Siamese neural network architecture with supplementary classification and regression branches

During the model training, various distributions of error function weights for supplementary branches were explored. Initially, the model utilized a uniform weight distribution, followed by adjustments where the additional outputs influenced the training outcomes by 50%, 20%, and 10% compared to the error function weights for the contrastive loss output. The refined Siamese neural network model reports an accuracy of 96.54%.

Following the training, it was observed that the supplementary branches had minimal impact on the model's accuracy, with a mean decrease of 1.11%.

Notably, the most commendable performance came from the model with two extra outputs, where their weight influence on the error function was at 50%.

However, the accuracy only rose by 0.72%, which might be attributed to a statistical anomaly. This presumption of statistical error is reinforced by the results from models with two extra outputs but with error function weights of 20% and 10% – these models exhibited accuracy variations of -0.34% and +0.36%, compared to the singular output model (Table 2). Such patterns indicate that, contrary to aiding the main model in honing in on distinct features, the extra branches seemingly degrade the training outcomes.

Table 2 – **Comparison of training results with different setups of supplementary tasks**

Additional outputs	Accuracy
No additional outputs	96.54%
Classification and regression, loss weights 100%	95,19%
Classification and regression, loss weights 50%	97,26%
Classification and regression, loss weights 20%	96,20%
Classification and regression, loss weights 10%	96,90%
Classification, loss weights 100%	91,00%
Classification, loss weights 50%	94,35%
Classification, loss weights 20%	95,43%
Classification, loss weights 10%	96,24%
Regression, loss weights 100%	93,92%
Regression, loss weights 50%	94,62%
Regression, loss weights 20%	96,61%
Regression, loss weights 10%	96,60%

Conclusions

The paper investigates the potential advantages of enhancing Siamese neural networks with supplementary outputs, drawing inspiration from the GoogLeNet model. Two primary datasets were employed: Fashion MNIST and a subset of the PlantVillage dataset.

The Siamese network's foundational architecture was juxtaposed against modified versions, including additional outputs targeting classification, regression, or both.

Experiments utilized the ResNet50V2 model for its efficiency and generalization capabilities, with results suggesting that while the supplementary outputs marginally influenced accuracy, the added complexity didn't always translate to superior performance. Notably, the model with dual outputs, where error function weights of the supplementary branches were set at 50%, yielded the most promising results. However, the overall impact of additional outputs on accuracy was inconsistent, highlighting the need for further exploration into their optimal integration.

REFERENCES

1. Lim, E.A., Wei, T. and Kadri, J.A. (2019), "A new formula to determine the optimal dataset size for training neural networks", *ARPN Journal of Engineering and Applied Sciences*, vol. 14, pp. 52-61, available at: http://www.arpnjournals.org/jeas/research_papers/rp_2019/jeas_0119_7525.pdf.
2. Hunt, B., Kwan, E., Dosdall, D., MacLeod R.S. and Ranjan, R. (2021), "Siamese Neural Networks for Small Dataset Classification of Electrograms", *2021 Computing in Cardiology (CinC)*, pp. 1-4, doi: 10.23919/CinC53138.2021.9662707.
3. Wiggers, K.L., Britto, A.S., Heutte, L., Koerich, A.L. and Oliveira, L.S. (2019), "Image Retrieval and Pattern Spotting using Siamese Neural Network", *2019 Int. Joint Conf. on NNw*, pp. 1–10, doi: https://doi.org/10.1109/IJCNN.2019.8852197
4. Guillerme, T., Puttick, M.N., Marcy, A.E. and Weisbecker, V. (2020), "Shifting spaces: Which disparity or dissimilarity measurement best summarize occupancy in multidimensional spaces?", *Ecology and Evolution*, vol. 10, pp.7761-7775, doi: https://doi.org/10.1002/ece3.6452.
5. Ivan, Dokmanic, Reza, Parhizkar, Juri, Ranieri and Martin, Vetterli (2015), "Euclidean Distance Matrices: Essential Theory, Algorithms and Applications", *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 12–30, doi: 10.1109/MSP.2015.2398954.
6. Malialisa, K., Panayiotoua, C.G. and Polycarpou, M.M. (2020), "Data-efficient Online Classification with Siamese Networks and Active Learning", *Int. Joint Conf. on Neural Networks (IJCNN)*, pp.1-7, doi: https://doi.org/10.48550/arXiv.2010.01659.
7. Lialin, V., Deshpande, V. and Rumshisky, A. (2023), "Scaling Down to Scale Up: A Guide to Parameter-Efficient Fine-Tuning", *arXiv preprint arXiv:2303.15647*, pp.5–12, available at: https://arxiv.org/abs/2303.15647.
8. Szegedy, C., Liu, W., Jia, Ya., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. and Rabinovich, A. (2014), "Going Deeper with Convolutions", *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, doi: https://doi.org/10.1109/CVPR.2015.7298594.
9. Han, Xiao, Kashif, Rasul and Roland, Vollgraf (2017), "Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms", *arXiv arXiv:1708.07747*, pp. 2–3, available at: https://arxiv.org/abs/1708.07747.
10. De, S., Mukherjee, A. and Ullah, E. (2018), "Convergence guarantees for RMSProp and ADAM in non-convex optimization and an empirical comparison to Nesterov acceleration", *arXiv:1807.06766*, pp. 5-10, available at: https://arxiv.org/abs/1807.06766.
11. Groenendijk, R., Karaoglu, S., Gevers, T. and Mensink, T. (2020), "Multi-Loss Weighting with Coefficient of Variations", *2021 IEEE Winter Conf. on Appl. of Comp. Vision*, pp. 1468-1477, doi: https://doi.org/10.1109/WACV48630.2021.00151.
12. Singh, D., Jain, N., Jain, P., Kayal, P., Kumawat, S. and Batra, N. (2019), "PlantDoc: A Dataset for Visual Plant Disease Detection", *Association for Computing Machinery*, pp. 249–253, doi: https://doi.org/10.1145/3371158.3371196.
13. Kaiming, He, Xiangyu, Zhang, Shaoqing, Ren and Jian, Sun (2016), "Identity Mappings in Deep Residual Networks", *Computer Vision – ECCV*, pp. 2–13, doi: https://doi.org/10.1007/978-3-319-46493-0_38.

Received (Надійшла) 14.07.2023
Accepted for publication (Прийнята до друку) 06.09.2023

ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

Мельниченко Артем Васильович – аспірант, асистент кафедри цифрових технологій в енергетиці, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського, Київ, Україна;
Artem Melnychenko – PhD Student, assistant of Department of Digital Technologies in Energy National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine;
e-mail: artemxl@gmail.com; ORCID ID: https://orcid.org/0009-0000-3588-4772.

Здор Костянтин Андрійович – аспірант, асистент кафедри цифрових технологій в енергетиці, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського, Київ, Україна;
Kostyantyn Zdor – PhD Student, assistant of Department of Digital Technologies in Energy National Technical University of Ukraine «Igor Sikorsky Kyiv Polytechnic Institute», Kyiv, Ukraine;
e-mail: kostya9919moonlight@gmail.com; ORCID ID: https://orcid.org/0009-0008-7640-1499.

Ефективність використання додаткових виходів у сіамських нейронних мережах

А. В. Мельниченко, К. А. Здор

Анотація. У галузі комп'ютерного зору ефективна обробка великої кількості зображень є комплексною задачею, яка вимагає використання глибоких нейронних мереж. Сіамські нейронні мережі, відомі своєю дзеркальною структурою, пропонують ефективне вирішення задач порівняння зображень, особливо обмеженого об'єму даних. У цьому дослідженні розглядається можливість покращення цих моделей шляхом додавання допоміжних виходів, які поліпшують точність класифікації і виявлення конкретних особливостей даних. В статті розглядається результати двох експериментів з використанням датасетів Fashion MNIST і PlantVillage, з включенням додаткової класифікації, регресії та комбінованих стратегій виходу з різними конфігураціями втрати ваги. Результати експериментів продемонстрували, що для простіших датасетів введення додаткових вихідних даних призводить до зниження точності моделі. І навпаки, для складніших датасетів оптимальна точність була досягнута за рахунок одночасної інтеграції додаткових виходів з регресією та класифікацією. Слід зазначити, що отримане підвищення точності є відносно незначним і не гарантує суттєвий вплив на загальну точність моделі.

Ключові слова: комп'ютерний зір; нейронні мережі; сіамські нейронні мережі; розпізнавання зображень.