ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2.
Available from http://www.i6doc.com/en/.

Antagonism between Classification and Reconstruction Processes in Deep Predictive Coding Networks

Jan Rathjens and Laurenz Wiskott

Ruhr University Bochum - Faculty of Computer Science
Universitaetsstr. 150 - 44801 Bochum - Germany

Abstract. Predictive coding-inspired deep networks for visual computing integrate classification and reconstruction processes in shared intermediate layers. Although synergy between these processes is commonly assumed, it has yet to be convincingly demonstrated. In this study, we utilize a purposefully designed family of autoencoder-like architectures with an added classification head to examine the consequences of combining classification- and reconstruction-driven information within the models' latent layers. Our findings underscore a significant challenge: Classification-driven information diminishes reconstruction-driven information in shared representations and vice versa. Our results challenge prevailing assumptions in predictive coding and offer guidance for future iterations of predictive coding concepts in deep networks.

1 Introduction

Predictive coding is a neuroscientific theory that postulates how sensory perception is performed in the brain, with a particular emphasis on visual information processing. According to this theory, each brain area predicts the activity of its preceding area along the visual pathway, transmitting these predictions through feedback connections in a top-down manner. Feedforward connections carry the difference between predicted and actual activity. Inference and learning are achieved by minimizing this difference.

Deep Predictive Coding Networks (DPCNs) are inspired by the principles of predictive coding and incorporate elements of this theory into artificial neural networks [4, 7, 3]. DPCNs are perceived as potentially transformative for predictive coding because of their scalability, efficiency, and modularity - properties absent in traditional computational models of predictive coding. Additionally, incorporating predictive coding concepts into deep learning models can potentially transfer desirable features of human visual processing, such as robustness to noise, sample efficiency, and generalization, to deep learning applications [2].

A common assertion among DPCNs is that integrating classification and reconstruction tasks into shared intermediate layers might synergistically enhance overall performance. However, this assumption faced scrutiny when Rane et al. [5] showed on the DPCN PredNet that incorporating a classification process into the architecture leads to a decline in the quality of the predicted images. Additionally, the classification accuracy remains inferior to that of comparable feedforward networks.

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

In this study, we explore whether the antagonism observed in PredNet is unique to its architecture or if such phenomena could also be observable in different DPCNs. Rather than examining other DPCN variants directly, our investigation is grounded on the premise that for DPCNs to achieve synergistic integration of these processes, their intermediate layers must effectively combine classification- and reconstruction-driven information synergistically. We, therefore, analyze the consequences of combining the two kinds of information into a shared intermediate layer in the context of deep learning by utilizing a purposefully designed model architecture.

Our results reveal a significant challenge in deep learning: Integrating classification and reconstruction-driven information into a shared representation is only possible to a certain extent. These findings challenge prevailing assumptions in DPCNs and point to potential future research directions to combine predictive coding principles with deep learning effectively.

2 Methods

To analyze the effects of integrating classification and reconstruction-driven information within a shared representation in deep learning, we utilize a purposefully designed model architecture we call Classification-Reconstruction Encoder (CRE); see Figure 1. The architecture is reminiscent of an autoencoder featuring an additional classification head connected to the latent $\boldsymbol{z}$-layer.

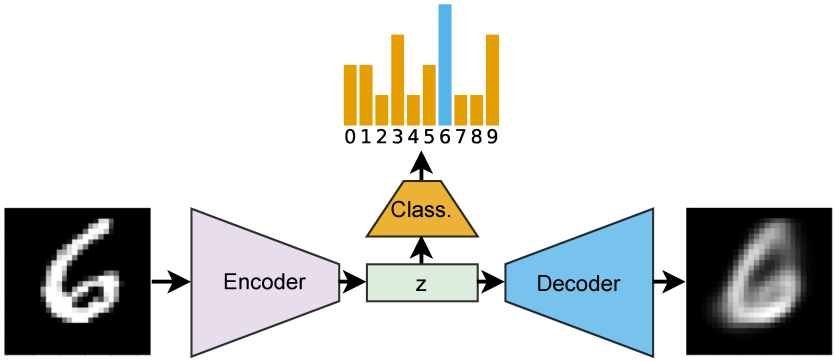


Fig. 1: Classification-Reconstruction Encoder (CRE) includes a decoder and classifier linked to the latent representation z. The encoder optimizes to encode input images for classification or reconstruction purposes.

Our model's encoder explicitly integrates classification and reconstruction-driven information within the shared $\mathbf{z}$-layer, with the classification and decoder heads only having access to this layer and being trained using cross-entropy loss L_{CE} and mean squared error loss L_{MSE}, respectively. The encoder is trained based on a weighted linear combination of both losses; see Equation (1). The λ-parameter adjusts the balance between the two types of information in $\mathbf{z}$.

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

A λ-value of 0 optimizes $\mathbf{z}$ for classification only, functioning like a traditional classification network, where the decoder is trained to reconstruct the image from a purely classification-driven representation. Conversely, a λ-value of 1 optimizes $\mathbf{z}$ solely for reconstruction, making the encoder-decoder setup function like a standard autoencoder, with the classifier working from a purely reconstruction-driven representation.

$$L = \lambda L_{\text{MSE}} + (1 - \lambda) L_{\text{CE}}, \quad \lambda \in [0, 1] \tag{1}$$

To ensure our findings are independent of specific model design choices, we base the CREs' components (encoder, decoder, classifier) on different common deep learning techniques. These variants include a fully connected (FC)-based CRE, a convolutional neural network (CNN)-based CRE, and a Vision Transformer (ViT)-based CRE [1]. We implement diverse configurations within these modules, e.g., we modify the dimensions of the $\boldsymbol{z}$-layer and adjust the complexity of the modules by altering their layer or parameter count. Detailed model descriptions are available in our GitHub repository[1]. We train these CREs on different datasets (MNIST, FashionMNIST, CIFAR-10) with different λ values, resulting in approximately 10,000 trained instances.

3 Results

Figure 2 showcases box plots illustrating the performance distributions across different λ-values for all three variants of the CRE with select configurations on various datasets. The plots reveal a consistent pattern. The models reach peak classification and reconstruction performances at the extreme values of λ. Specifically, $\lambda = 0$ yields the best classification results, while $\lambda = 1$ leads to the most effective reconstruction. On the other hand, these extreme λ-values also correspond to the lowest performances for the opposite metric - poor reconstruction at $\lambda = 0$ and suboptimal classification at $\lambda = 1$. We also note a mostly monotonic relationship between λ-values and performance: As λ increases, reconstruction performance improves, whereas classification performance deteriorates. The reconstruction performance of the FC-based CRE on the MNIST dataset slightly deviates from this relationship as the pattern is less consistent than in the other plots. The consistent pattern across architectures and datasets highlights a general trade-off between integrating classification and reconstruction information: Enhancing one aspect tends to weaken the other without synergistic gains. This trade-off is especially evident when transferring representations from one task to another, such as using purely reconstruction-driven representations for classification purposes and vice versa.

We exemplary visualize the latent space of FC-based CREs with a three-dimensional $\mathbf{z}$-layer and sample reconstructions for distinct λ-values on FashionMNIST in Figure 3. For the latent spaces, at $\lambda = 0.0$, classes form nearly straight lines radiating from a central point in a star-like pattern. With $\lambda = 0.2$,

[1] `https://github.com/wiskott-lab/classification-reconstruction-encoder`

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

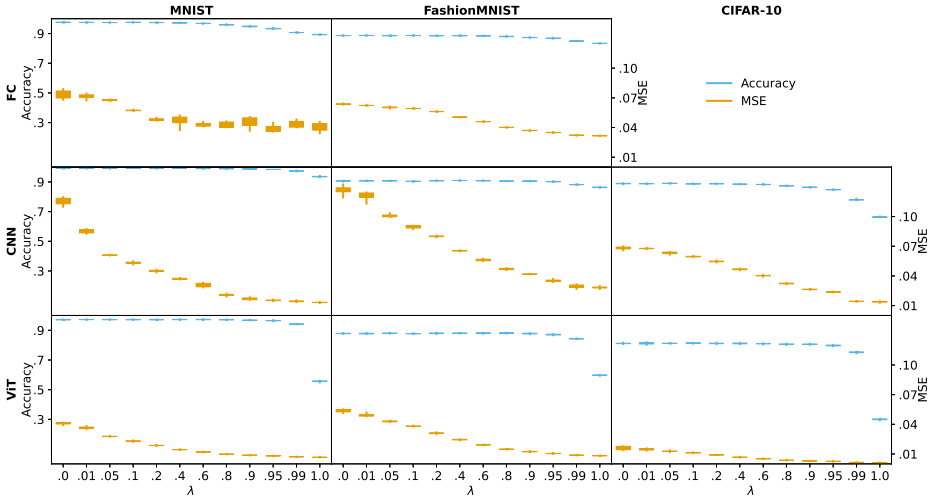
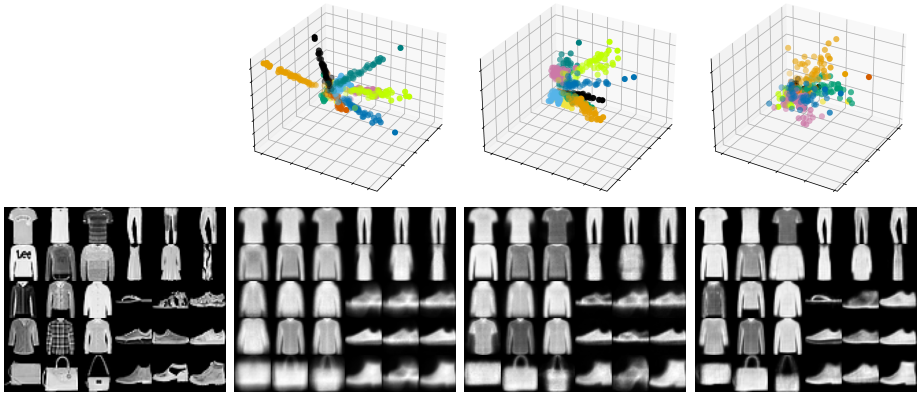


Fig. 2: Performance w.r.t. λ. The box plots show classification (blue) and reconstruction (orange) performance across exemplary CRE variants and datasets.

this radial layout persists but with increased scatter among points. At $\lambda = 1.0$, the arrangement shifts markedly, with classes appearing as overlapping clusters rather than linear extensions. For the reconstructions, at $\lambda = 0$, images are similar across classes, producing prototypical images without distinct input details. With $\lambda = 0.2$, reconstructions include some details from the inputs, blending prototypical and detailed features. At $\lambda = 1.0$, there is a marked improvement in detail capture in the reconstructions. Visualizing the latent space and sampling reconstructions offers insight into the trade-off effect. Classification-driven approaches favor a star-shaped configuration in a three-dimensional latent space, whereas reconstruction-driven approaches prefer point clouds. Combining these preferences leads to compromises, such as less distinct lines or less diverse point clouds, a trend likely extending to higher dimensions. Moreover, integrating classification-driven information often results in less visual detail, as decoders tend to reconstruct based on prototypical examples, losing finer details.

Figure 4 illustrates the trade-off effect with respect to the number of latent dimensions in FC-based CREs for MNIST and FashionMNIST, comparing various λ-values against Principal Component Analysis (PCA)- and Random Projection (RP)-based representations. At each dimension, the trade-off effect is reflected in the distance to the best-performing λ-value (0 for classification, 1 for reconstruction). A larger gap indicates a significant trade-off, while a smaller gap suggests a modest impact. For MNIST and FashionMNIST, classification performance generally improves with increasing dimensions, but $\lambda = 0$ reaches peak performance at lower dimensions (e.g., $\dim(\boldsymbol{z}) = 4$) with minimal improvement beyond. As dimensions increase, the performance gap to $\lambda = 0$ narrows, with all λ-values under one nearly closing this gap at higher dimensions, though

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.



(a) Input images (b) $\lambda = 0.0$ (c) $\lambda = 0.2$ (d) $\lambda = 1.0$

Fig. 3: Visualization of latent spaces and reconstructions. Top row: Latent spaces using 3D plots from an exemplary FC-based CRE with a 3-dimensional latent space for each λ trained on FashionMNIST. Each plot shows fifty instances per class, each represented by a different color. Bottom row: Exemplary reconstructions for the same models.

less so for $\lambda = 1$, PCA-, and RP-based representations. Regarding reconstruction, efficiency consistently improves with dimension regardless of representation type, with variability in performance ranking among types. Low-dimension best reconstructions are often at $\lambda = 1$ or 0.99, but PCA can match or exceed these at higher dimensions. RP-based representations perform worse in classification but improve in reconstruction at higher dimensions, sometimes surpassing classification-focused representations. Our analysis shows that increasing the latent space size effectively mitigates the trade-off effect, suggesting a competition for resources between classification- and reconstruction-driven representations.

4 Conclusion

In this study, we explored how classification and reconstruction-driven information interact within shared representations in deep neural networks. Our findings reveal a consistent trade-off: neither process benefits from including the other. Increasing the dimensions of shared representations mitigates this effect, indicating a competition for resources. These results prompt not only a reevaluation of current DPCNs but also classic computational models of predictive coding, which face difficulties in achieving satisfactory classification and reconstruction performance simultaneously [6]. Our study's limitations include a focus on a single layer, rather than an integration across multiple hierarchical layers. Moreover, despite employing a broad array of deep learning modules, we cannot preclude the existence of a module that effectively integrates both types of information, meriting further investigations.

ESANN 2024 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 9-11 October 2024, i6doc.com publ., ISBN 978-2-87587-090-2. Available from http://www.i6doc.com/en/.

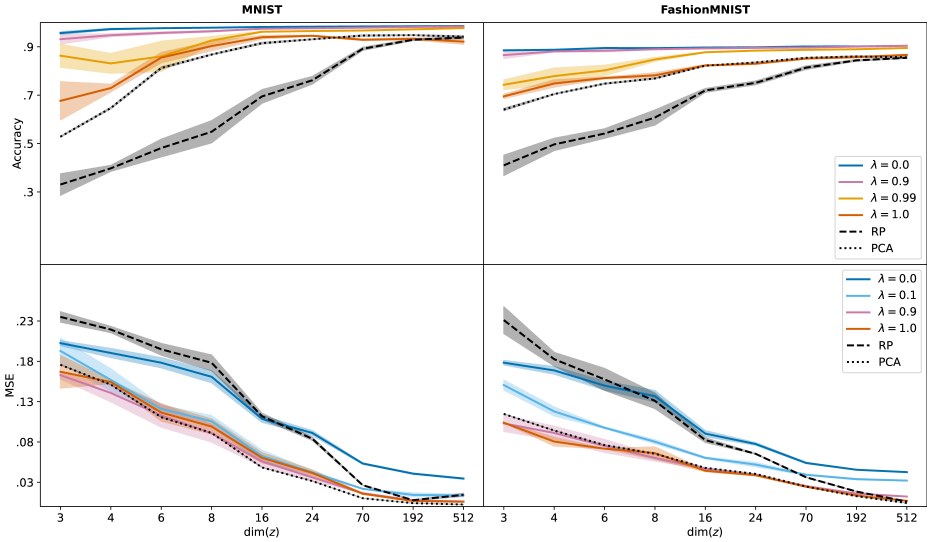


Fig. 4: Trade-off by latent space dimensionality: Plots show average classification (top) and reconstruction (bottom) performances of FC-based CREs on MNIST and FashionMNIST across λ-values, with a 95% confidence interval. Results for RP-based and PCA-based representations are included.

References

[1] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *International Conference on Learning Representations*, Oct. 2020.

[2] R. Geirhos, D. H. J. Janssen, H. H. Schütt, J. Rauber, M. Bethge, and F. A. Wichmann. Comparing deep neural networks against humans: object recognition when the signal gets weaker, Dec. 2018. arXiv:1706.06969 [cs, q-bio, stat].

[3] Y. Huang, J. Gornet, S. Dai, Z. Yu, T. Nguyen, D. Tsao, and A. Anandkumar. Neural Networks with Recurrent Generative Feedback. In *Advances in Neural Information Processing Systems*, volume 33, pages 535–545. Curran Associates, Inc., 2020.

[4] W. Lotter, G. Kreiman, and D. D. Cox. Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings*. OpenReview.net, 2017.

[5] R. P. Rane, E. Szügyi, V. Saxena, A. Ofner, and S. Stober. PredNet and Predictive Coding: A Critical Review. In C. Gurrin, B. P. Jónsson, N. Kando, K. Schöffmann, Y.-P. P. Chen, and N. E. O'Connor, editors, *Proceedings of the 2020 on International Conference on Multimedia Retrieval, ICMR 2020, Dublin, Ireland, June 8-11, 2020*, pages 233–241. ACM, 2020.

[6] W. Sun and J. Orchard. A Predictive-Coding Network That Is Both Discriminative and Generative. *Neural Computation*, 32(10):1836–1862, Oct. 2020.

[7] H. Wen, K. Han, J. Shi, Y. Zhang, E. Culurciello, and Z. Liu. Deep Predictive Coding Network for Object Recognition. In *Proceedings of the 35th International Conference on Machine Learning*, pages 5266–5275. PMLR, July 2018. ISSN: 2640-3498.