A more selective integration function to improve deep neural network models of visual perception

Michael W. Spratling (`michael.spratling@uni.lu`) and **Heiko H. Schütt** (`heiko.schutt@uni.lu`)
University of Luxembourg, Department of Behavioural and Cognitive Sciences, L-4366 Esch-sur-Alzette, Luxembourg.

Abstract

Human visual perception remains substantially more robust than computer vision. We hypothesised that this might be due to the higher selectivity of biological neurons compared to artificial neurons which typically employ a linear integration function that is poor at feature detection. To test this hypothesis we replaced the convolutional layers in deep neural networks (DNNs) with a new integration function, the Consistent Intensity Metric (CIM). We trained networks based on CIM on six benchmark image classification tasks (MNIST, FashionMNIST, SVHN, CIFAR10, CIFAR100, and TinyImageNet) and compared the performance of these networks with equivalent convolutional neural networks matched to have an equal number of parameters. Consistent with our hypothesis, the CIM-based networks were better able to generalise from the training data. This was demonstrated by higher accuracy on both the standard test data and distorted input images (the common corruptions data-sets). Furthermore, test images that did not belong to any of the categories in the training data-set were less likely to be misclassified as belonging to one of the known categories. Our results suggest that using a more selective integration function can help address some of the reliability and robustness issues of DNNs. As these issues do not affect humans, this modification also makes DNNs functionally more similar to the biological visual system.

Keywords: Image Classification; Generalisation; Out-of-Distribution Rejection; Integration Functions; Pooling Functions; Loss Functions

Since the inception of neural networks (NNs) over 65 years ago (Rosenblatt, 1958) in almost all multi-unit models the neurons combine their inputs using a weighted sum. This is implemented using either matrix multiplication (in fully-connected layers) or cross-correlation (in convolutional layers). Such linear integration is used in "linear-nonlinear" models, "integrate-and-fire" models, multi-layer perceptrons, convolutional neural networks (CNNs), transformers (where it is used to calculate the queries, keys and values), generative models, predictive coding models, recurrent NNs, LSTMs (where it is used to calculate the "input", "forget", and "output" values), and many others.

Research in both machine learning and computational neuroscience has, therefore, tended to preserve the linear integration function while trying to improve the accuracy of models by modifying other factors such as the activation function: the non-linear operation applied to the output of the integration function. This has resulted in a large variety of activation functions (sigmoid, tanh, ReLU, GELU, ELU, PReLU, SiLU, Mish, Swish, Softplus, Softmax, *etc.*), as well as multivariate activation functions that operate on the outputs of multiple linear integration functions (as in LSTMs, transformers, divisive normalisation, and some models cortical pyramidal cells (Phillips et al., 2024)). However, the linear integration function has poor selectivity, as illustrated in Fig. 1. It is highly sensitive to the average intensity of the pre-synaptic activity but relatively insensitive to the configuration of those inputs. Therefore, the ubiquity of linear integration is odd given that biological neurons have been (and still are) often characterised as highly specific feature detectors (Barlow, 1953).

Table 1: A comparison of the performance of CNNs composed of standard building-blocks ('std') with equivalent CIM-based networks composed of the proposed building-blocks ('our'). Bold text indicates the best performance on each metric for each training data-set (all metrics range from 0 to 100 with higher values being better). OOD rejection performed using Maximum Softmax Probability (Hendrycks and Gimpel, 2017) as the prediction confidence score.

	Num. Params.	Training Data-set	Clean Accuracy (%)	Corrupt Accuracy (%)	OOD rejection AUROC (%)
std	44170	MNIST	98.75 ± 0.06	71.46 ± 1.62	95.37 ± 0.97
our	44160	MNIST	$\mathbf{99.10 \pm 0.10}$	$\mathbf{84.68 \pm 1.43}$	$\mathbf{97.98 \pm 0.47}$
std	84916	FMNIST	90.55 ± 0.31	44.69 ± 0.67	78.55 ± 3.43
our	84240	FMNIST	$\mathbf{91.22 \pm 0.05}$	$\mathbf{54.01 \pm 2.40}$	$\mathbf{97.86 \pm 0.36}$
std	266954	SVHN	89.80 ± 0.37	$\mathbf{51.41 \pm 0.74}$	85.70 ± 0.53
our	264400	SVHN	$\mathbf{93.88 \pm 0.43}$	45.96 ± 2.20	$\mathbf{98.00 \pm 0.32}$
std	458218	CIFAR10	83.39 ± 0.23	71.01 ± 0.23	72.34 ± 2.52
our	459728	CIFAR10	$\mathbf{86.57 \pm 0.16}$	$\mathbf{71.74 \pm 0.67}$	$\mathbf{82.17 \pm 2.74}$
std	676730	CIFAR100	$\mathbf{62.90 \pm 0.35}$	39.53 ± 0.19	63.77 ± 1.07
our	679296	CIFAR100	62.66 ± 0.38	$\mathbf{40.66 \pm 0.17}$	$\mathbf{73.56 \pm 4.20}$
std	2.15M	TinyIN	39.25 ± 0.23	10.93 ± 0.23	51.17 ± 2.02
our	2.15M	TinyIN	$\mathbf{44.07 \pm 0.40}$	$\mathbf{12.75 \pm 0.44}$	$\mathbf{53.37 \pm 4.99}$
	mean diff. (our-std) over data-sets		+2.14	+3.46	+9.34

We propose a new neural response function, the Consistent Intensity Metric (CIM), that is more selective for the configuration of inputs while also being relatively robust to changes in appearance (Fig. 1 last column). CIM is illustrated in Fig. 2. CIM measures the ratio between the synaptic inputs and the corresponding synaptic weights and produces an activation that is inversely proportional to the variance of these ratios. A strong response, therefore, requires that the intensity of the inputs across the receptive field (RF) is consistent with the

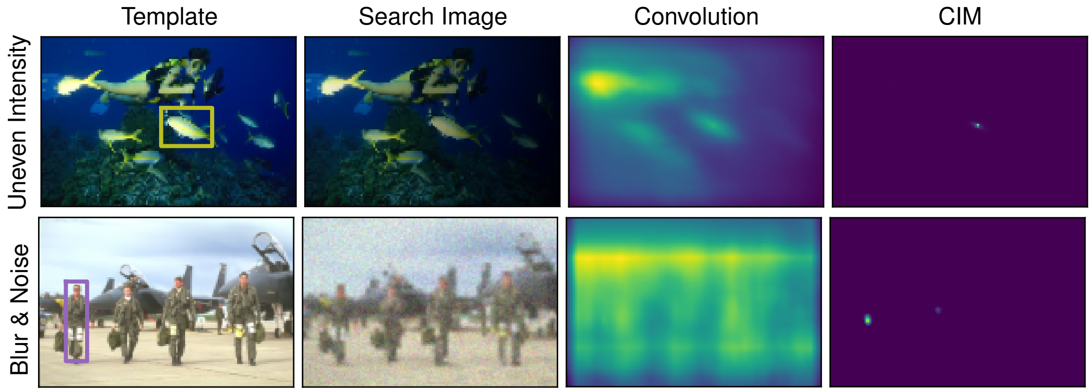

Figure 1: Comparison of linear integration and CIM on a template matching task. The columns show: 1) the original image and a bounding box indicating the location where the template was extracted; 2) the image to which the template was compared; the similarity between the template and the searched image calculated by 3) linear filtering, and 4) CIM. In the first row the search image was produced by non-uniformly changing the intensity of the original. In the second row the search image was produced by blurring followed by the addition of uniform random noise. Images come from the BSDS500 data-set (Martin et al., 2001).

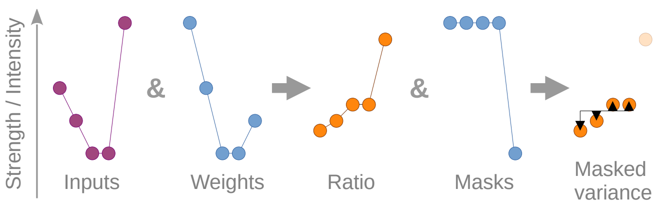

Figure 2: The calculation performed by the proposed integration function (CIM). A neuron with five inputs is illustrated, and the inputs and the weights are shown in the first two panels. Instead of multiplying these corresponding values together, as in the standard integration function, the ratio between the weights and corresponding inputs is calculated. The output of the neuron is inversely related to the variance of these ratios. Each neuron also learns mask values associated with each input. Masks are used to weigh the contribution of different inputs to the variance, enabling certain inputs to be ignored as shown in the last panel.

pattern stored in the weights. Each neuron also learns a second parameter associated with each synapse (the masks) that allows it to learn the importance of each input to the feature it represents: *i.e.*, to allow each neuron to learn to refine its RF.

We replaced the standard convolutional layers and ReLU activation functions in fully-convolutional NNs with CIM. However, for such networks to produce performance better than conventional CNNs it was also found necessary to replace other standard CNN building-blocks. Specifically, Max pooling was replaced with Boltzmann pooling. Boltzmann pooling uses a smooth approximation to the maximum function to integrate values within a spatial region of the input feature-map and (in contrast to standard max pooling) over a number of input channels. Cross-Entropy loss was replaced with High Error Margin (HEM) loss (Spratling and Schütt, 2025). HEM is a variant of multi-class margin (MM) loss (Crammer and Singer, 2002) that concentrates on minimising the most confident misclassifications. We compared the performance of standard and CIM-based architectures that were identical (same depth, kernel dimensions, spatial pooling size, padding) except for the changes to the building-blocks described above and number of output channels in each layer. The latter was varied to ensure that the networks being compared had trainable parameter numbers as similar as possible.

CIM-based models perform better than equivalent standard CNNs on a number of image classification benchmarks in terms of clean accuracy, generalisation to image corruption, and out-of-distribution (OOD) rejection (Table 1). Despite CIM being designed to be more selective it shows a greater ability to generalise: CIM-based models more accurately classify samples from both the standard test sets and the common-corruptions data-sets. The increase in performance in classifying samples from known classes is accompanied by a consistent increase in the ability to distinguish known from unknown image categories.

Our results strongly suggest that weighted summation may not be the optimal integration function for modelling visual perception. In future we aim to more thoroughly evaluate the generalisation and robustness of CIM-based networks (Croce et al., 2021; Geirhos et al., 2021; Li et al., 2024), their interpretability (Doshi-Velez and Kim, 2017), and their ability to account for biological visual processing (Conwell et al., 2024; Geirhos et al., 2018; Schrimpf et al., 2020).

Acknowledgements

The simulations were performed using the Luxembourg national supercomputer MeluXina and the HPC facilities of the University of Luxembourg (Varrette et al., 2022). The authors gratefully acknowledge the LuxProvide and University of Luxembourg teams for their expert support.

References

Barlow, H. B. (1953). Summation and inhibition in the frog's retina. *Journal of Physiology*, 119:69–88. doi:10.1113/jphysiol.1953.sp004829.

Conwell, C., Prince, J. S., Kay, K. N., Alvarez, G. A., and Konkle, T. (2024). A large-scale examination of inductive biases shaping high-level visual representation in brains and machines. *Nature Communications*, 15(9383). doi:10.1038/s41467-024-53147-y.

Crammer, K. and Singer, Y. (2002). On the algorithmic implementation of multiclass kernel-based vector machines. *Journal of Machine Learning Research*, 2:265–92.

Croce, F., Andriushchenko, M., Sehwag, V., Debenedetti, E., Flammarion, N., Chiang, M., Mittal, P., and Hein, M. (2021). Robustbench: a standardized adversarial robustness benchmark. In *Proceedings of the Conference on Advances in Neural Information Processing Systems*. `https://openreview.net/forum?id=SSKZPJCt7B`.

Doshi-Velez, F. and Kim, B. (2017). Towards A rigorous science of interpretable machine learning. arXiv:1702.08608.

Geirhos, R., Narayanappa, K., Mitzkus, B., Thieringer, T., Bethge, M., Wichmann, F. A., and Brendel, W. (2021). Partial success in closing the gap between human and machine vision. In *Proceedings of the Conference on Advances in Neural Information Processing Systems*. arXiv:2106.07411.

Geirhos, R., Temme, C. R. M., Rauber, J., Schütt, H. H., Bethge, M., and Wichmann, F. A. (2018). Generalisation in humans and deep neural networks. In *Proceedings of the Conference on Advances in Neural Information Processing Systems*. arXiv:1808.08750.

Hendrycks, D. and Gimpel, K. (2017). A baseline for detecting misclassified and out-of-distribution examples in neural networks. In *Proceedings of the International Conference on Learning Representations*. `https://openreview.net/forum?id=Hkg4TI9xl`. arXiv:1610.02136.

Li, L., Wang, Y., Sitawarin, C., and Spratling, M. W. (2024). OODRobustBench: a benchmark and large-scale analysis of adversarial robustness under distribution shift. In *Proceedings of the International Conference on Machine Learning*. arXiv:2310.12793.

Martin, D., Fowlkes, C., Tal, D., and Malik, J. (2001). A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings of the International Conference on Computer Vision*, volume 2, pages 416–23.

Phillips, W. A., Bachmann, T., Spratling, M. W., Muckli, L., Petro, L. S., and Zolnik, T. (2024). Cellular psychology: relating cognition to context-sensitive pyramidal cells. *Trends in Cognitive Sciences*, 29(1):28–40.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386–408.

Schrimpf, M., Kubilius, J., Lee, M. J., Murty, N. A. R., Ajemian, R., and DiCarlo, J. J. (2020). Integrative benchmarking to advance neurally mechanistic models of human intelligence. *Neuron*, 108(3):413–23. doi:10.1016/j.neuron.2020.07.040.

Spratling, M. W. and Schütt, H. H. (2025). A margin-based replacement for cross-entropy loss. arXiv:2501.12191.

Varrette, S., Cartiaux, H., Peter, S., Kieffer, E., Valette, T., and Olloh, A. (2022). Management of an Academic HPC & Research Computing Facility: The ULHPC Experience 2.0. In *Proc. of the 6th ACM High Performance Computing and Cluster Technologies Conf. (HPCCT 2022)*, Fuzhou, China. Association for Computing Machinery (ACM).