

Chapter 1

Background

The foundational principle of neural networks is, in its purest form, inspired by the biology of the human brain. The field of AI has often modelled new algorithms after biological phenomena; in this field, genetic algorithms are based on evolution and particle swarm optimization is based on social behaviors.

LeCun et al's seminal work in this field, *Gradient-Based Learning Applied to Document Recognition* [9], provided the first basis of using backpropagation methodologies to train visual classifiers. Even more importantly, it introduced the fundamental structure of the modern visual deep learning network. In its usage of convolutions as a method for extracting high-level features out of larger images, it set the framework for a new style of network that would prove to be far more efficient and scalable.

A convolution is an operator applied to two functions f and g , which provides a way of interpreting one function in the context of the other. The operation is generally defined as

$$(f * g)(t) = \int_{-\infty}^{\infty} f(r)g(t - r)dr$$

In the perspective of modern deep learning, we are primarily interested in its usage as a matrix operator; in this context, we limit the range of g to the size of the matrix s such that

$$(f * g)(t) = \int_0^s f(r)g(t - r)dr$$

In this context, we refer to g as the *convolutional kernel*. Using a convolutional kernel to preprocess the image proves to be a critical to the performance of modern deep learning methods, as a small kernel can operate over a large image in parallel. Figure 1.1 shows a simple convolutional kernel can produce high-level details of an image. In this example, we consider the basic edge-detecting matrix

$$E = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Convolutional neural networks are therefore the product of using these

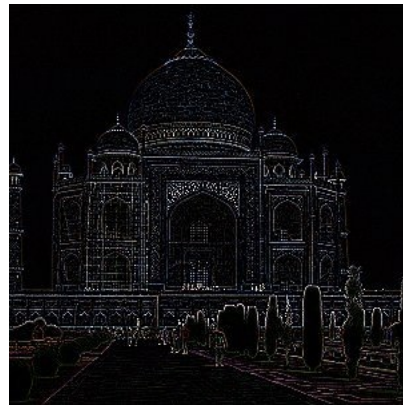


Figure 1.1: A demonstration of an edge-detecting convolution, from the GIMP User's Manual. [3]

Chapter 2

Related Works

In this work, we are primarily interested in optimizing neural network architectures and other hyperparameters.

Bibliography

- [1] CHANGPINYO, S., SANDLER, M., AND ZHMOGINOV, A. The power of sparsity in convolutional neural networks. *arXiv preprint arXiv:1702.06257* (2017).
- [2] COURBARIAUX, M., AND BENGIO, Y. Binarynet: Training deep neural networks with weights and activations constrained to +1 or -1. *arXiv preprint arXiv:1602.02830* (2016).
- [3] GIMP. Convolution matrix. <https://docs.gimp.org/en/plugin-convmatrix.html>.
- [4] HAN, S., POOL, J., TRAN, J., AND DALLY, W. Learning both weights and connections for efficient neural network. In *Advances in Neural Information Processing Systems* (2015), pp. 1135–1143.
- [5] HU, H., PENG, R., TAI, Y.-W., AND TANG, C.-K. Network trimming: A data-driven neuron pruning approach towards efficient deep architectures. *arXiv preprint arXiv:1607.03250* (2016).
- [6] HUANG, G., SUN, Y., LIU, Z., SEDRA, D., AND WEINBERGER, K. Q. Deep networks with stochastic depth. In *European Conference on Computer Vision* (2016), Springer, pp. 646–661.
- [7] IANDOLA, F. N., MOSKEWICZ, M. W., ASHRAF, K., HAN, S., DALLY, W. J., AND KEUTZER, K. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 1mb model size. *arXiv preprint arXiv:1602.07360* (2016).
- [8] LEBEDEV, V., AND LEMPITSKY, V. Fast convnets using group-wise brain damage. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 2554–2564.
- [9] LECUN, Y., BOTTOU, L., BENGIO, Y., AND HAFFNER, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86, 11 (1998), 2278–2324.
- [10] LECUN, Y., DENKER, J. S., SOLLA, S. A., HOWARD, R. E., AND JACKEL, L. D. Optimal brain damage. In *NIPS* (1989), vol. 2, pp. 598–605.
- [11] MURRAY, K., AND CHIANG, D. Auto-sizing neural networks: With applications to n-gram language models. *arXiv preprint arXiv:1508.05051* (2015).