Otto Mättas
245867IAXD

Scoping Good and Bad Figures

This assignment explores figure design in AI research by examining figures from seminal papers. This analysis categorises three "good" figures (presented as the first three figures), characterised by clarity and effective layout, and three "bad" figures (presented as the last three figures), where clutter or lack of detail reduces interpretability.

References

1. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). Attention Is All You Need. In Advances in Neural Information Processing Systems 30 (NeurIPS 2017). DOI: https://doi.org/10.48550/arXiv.1706.03762.
2. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In J. Burstein, C. Doran, & T. Solorio (Eds.), Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers) (pp. 4171-4186). Association for Computational Linguistics. DOI: https://doi.org/10.18653/v1/N19-1423.
3. Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Networks. In Advances in Neural Information Processing Systems 27 (NeurIPS 2014). DOI: https://doi.org/10.48550/arXiv.1406.2661.
4. He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016) (pp. 770-778). DOI: https://doi.org/10.48550/arXiv.1512.03385.
5. Bahdanau, D., Cho, K., & Bengio, Y. (2016). Neural Machine Translation by Jointly Learning to Align and Translate. In International Conference on Learning Representations (ICLR 2015). DOI: https://doi.org/10.48550/arXiv.1409.0473.
6. Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to Sequence Learning with Neural Networks. In Advances in Neural Information Processing Systems 27 (NeurIPS 2014). DOI: https://doi.org/10.48550/arXiv.1409.3215.

Attention Is All You Need

DOI: https://doi.org/10.48550/arXiv.1706.03762

Why Good: This figure communicates complex architectural details through well-organised layers and clear visual flow, with good use of colour coding to distinguish different components. The mathematical annotations and labelled arrows make the attention mechanism understandable while maintaining a clean appearance.

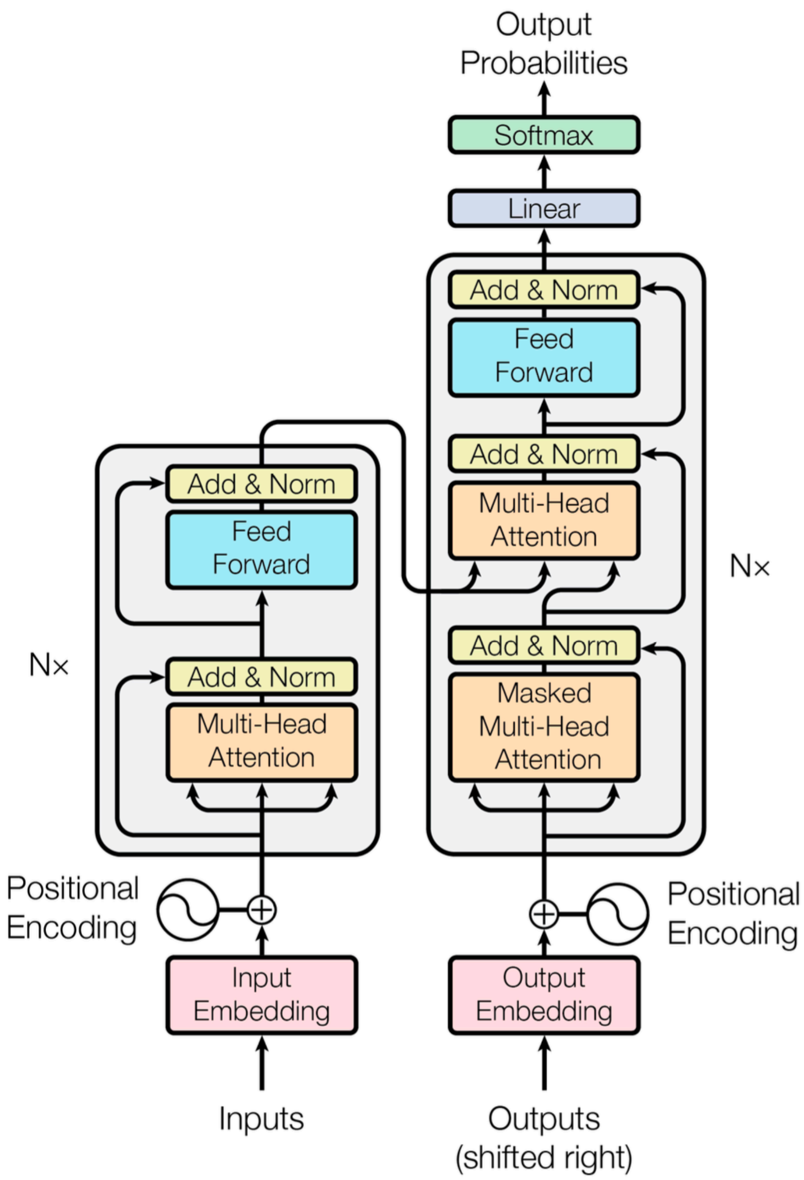

Figure 1: The Transformer - model architecture.

BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

DOI: https://doi.org/10.18653/v1/N19-1423

Why Good: The figure brilliantly illustrates the two-stage process of pre-training and fine-tuning through a logical, well-structured layout that guides the reader. The consistent use of visual elements and clear examples demonstrates the model's structure and its practical application, making abstract concepts understandable.

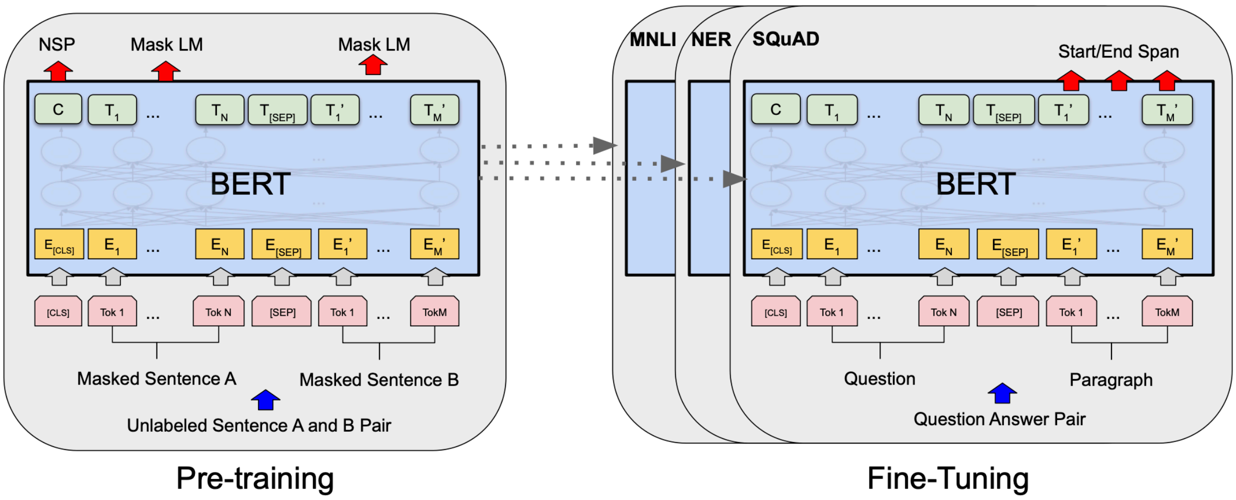

Figure 1: Overall pre-training and fine-tuning procedures for BERT. Apart from output layers, the same architectures are used in both pre-training and fine-tuning. The same pre-trained model parameters are used to initialize models for different down-stream tasks. During fine-tuning, all parameters are fine-tuned. `[CLS]` is a special symbol added in front of every input example, and `[SEP]` is a special separator token (e.g. separating questions/answers).

Generative Adversarial Nets

DOI: https://doi.org/10.48550/arXiv.1406.2661

Why Good: The figure achieves clarity through minimalist design, using simple geometric shapes and arrows to convey the adversarial relationship between components. The illustration of data flow and transformation processes helps readers grasp the sophisticated training dynamics without unnecessary complexity.

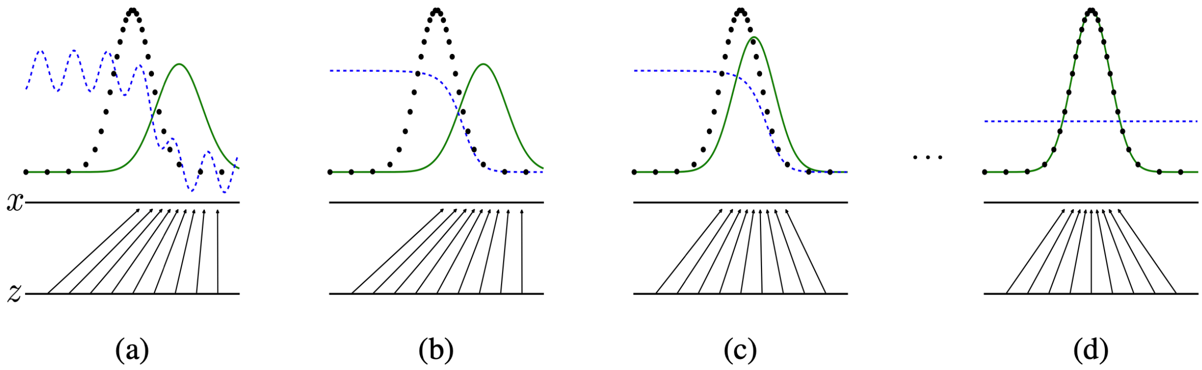

Figure 1: Generative adversarial nets are trained by simultaneously updating the **d**iscriminative distribution (D, blue, dashed line) so that it discriminates between samples from the data generating distribution (black, dotted line) $p_{\boldsymbol{x}}$ from those of the **g**enerative distribution p_g (G) (green, solid line). The lower horizontal line is the domain from which $\boldsymbol{z}$ is sampled, in this case uniformly. The horizontal line above is part of the domain of $\boldsymbol{x}$. The upward arrows show how the mapping $\boldsymbol{x} = G(\boldsymbol{z})$ imposes the non-uniform distribution p_g on transformed samples. G contracts in regions of high density and expands in regions of low density of p_g. (a) Consider an adversarial pair near convergence: p_g is similar to p_{data} and D is a partially accurate classifier. (b) In the inner loop of the algorithm D is trained to discriminate samples from data, converging to $D^*(\boldsymbol{x}) = \frac{p_{\text{data}}(\boldsymbol{x})}{p_{\text{data}}(\boldsymbol{x})+p_g(\boldsymbol{x})}$. (c) After an update to G, gradient of D has guided $G(\boldsymbol{z})$ to flow to regions that are more likely to be classified as data. (d) After several steps of training, if G and D have enough capacity, they will reach a point at which both cannot improve because $p_g = p_{\text{data}}$. The discriminator is unable to differentiate between the two distributions, i.e. $D(\boldsymbol{x}) = \frac{1}{2}$.

Deep Residual Learning for Image Recognition

DOI: https://doi.org/10.48550/arXiv.1512.03385

Recommendation: The figure requires better spacing between plot lines and larger font sizes to improve readability of performance metrics across network depths. The overlapping elements and compressed layout make it challenging to distinguish between different experimental conditions, which could be resolved by separating the data into multiple panels with consistent scales.

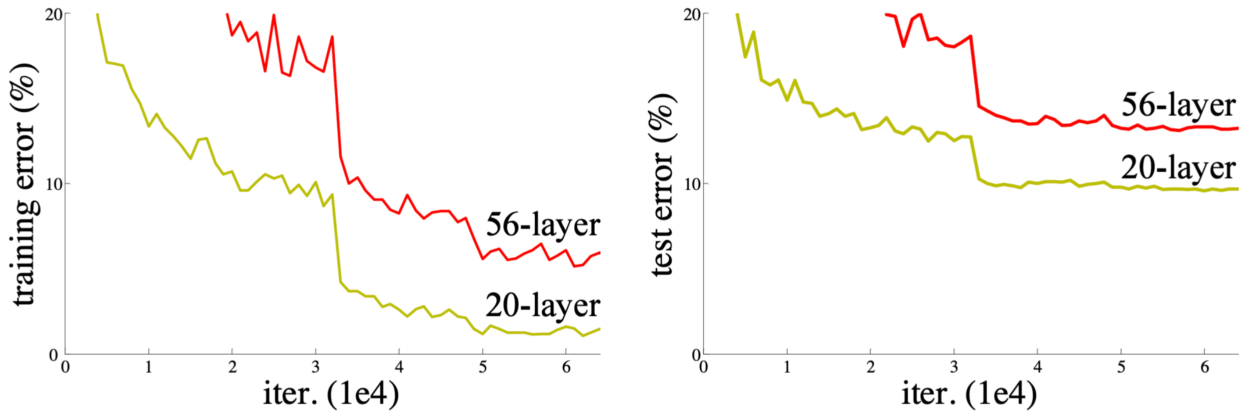

Figure 1. Training error (left) and test error (right) on CIFAR-10 with 20-layer and 56-layer "plain" networks. The deeper network has higher training error, and thus test error. Similar phenomena on ImageNet is presented in Fig. 4.

Neural Machine Translation by Jointly Learning to Align and Translate

DOI: https://doi.org/10.48550/arXiv.1409.0473

Recommendation: The figure needs a proper legend to explain the meaning of various symbols and connection types used in the alignment mechanism. The visual hierarchy could be enhanced through consistent use of symbols and clear differentiation between attention weights and neural network components.

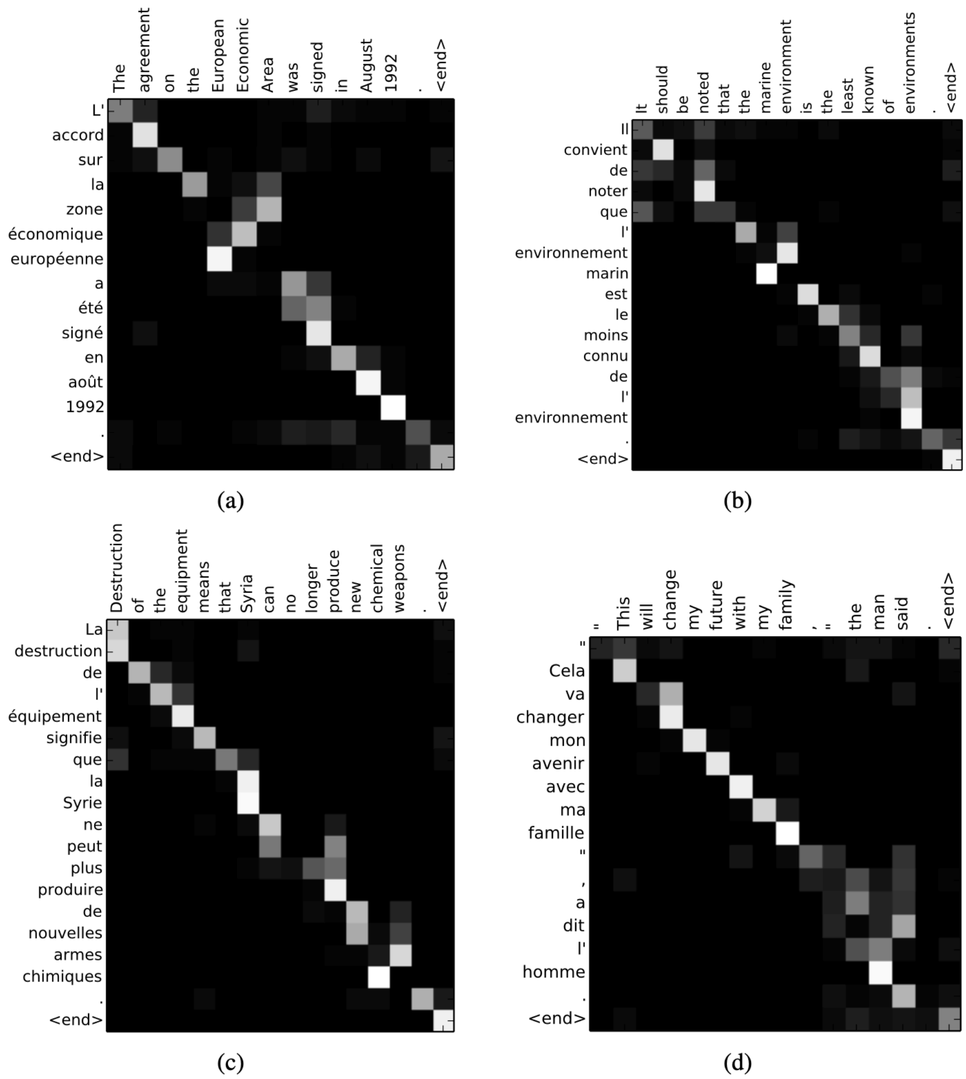

Figure 3: Four sample alignments found by RNNsearch-50. The x-axis and y-axis of each plot correspond to the words in the source sentence (English) and the generated translation (French), respectively. Each pixel shows the weight α_{ij} of the annotation of the j-th source word for the i-th target word (see Eq. (6)), in grayscale (0: black, 1: white). (a) an arbitrary sentence. (b–d) three randomly selected samples among the sentences without any unknown words and of length between 10 and 20 words from the test set.

Sequence to Sequence Learning with Neural Networks

DOI: https://doi.org/10.48550/arXiv.1409.3215

Recommendation: The figure would benefit from increased spacing between components and clearer directional indicators to show the temporal flow of information. A better organised layout with distinct encoder and decoder sections would help readers understand the sequential nature of the model's operation.

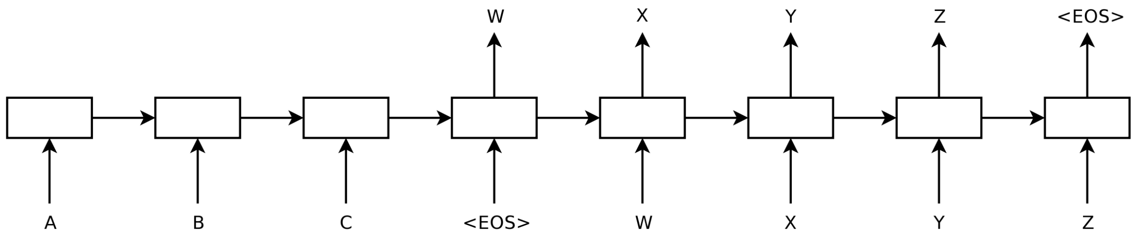

Figure 1: Our model reads an input sentence "ABC" and produces "WXYZ" as the output sentence. The model stops making predictions after outputting the end-of-sentence token. Note that the LSTM reads the input sentence in reverse, because doing so introduces many short term dependencies in the data that make the optimization problem much easier.