Adversarial Attacks on Images

Pierre-Gabriel Berlureau   Antoine Groudiev   Matéo Torrents

18th December 2024

1 Introduction

References

[1] Ian J Goodfellow, Jonathon Shlens and Christian Szegedy. 'Explaining and harnessing adversarial examples'. In: *arXiv preprint arXiv:1412.6572* (2014).

[2] Alexey Kurakin, Ian J Goodfellow and Samy Bengio. 'Adversarial examples in the physical world'. In: *Artificial intelligence safety and security*. Chapman and Hall/CRC, 2018, pp. 99–112.

[3] Seyed-Mohsen Moosavi-Dezfooli, Alhussein Fawzi and Pascal Frossard. 'Deepfool: a simple and accurate method to fool deep neural networks'. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2574–2582.