

**Feature Extraction Block**

Nowadays, many feature extraction blocks have been proposed. The main idea of the inception block [2] (Fig. 1.(c)) is to ﬁnd out how an optimal local sparse structure works in a convolutional network. However, these diﬀerent scale features simply concatenate together, which leads to the underutilization of local features. In 2016, Kim et al. [3] proposed a residual learning framework (Fig. 1.(a)) to ease the training of networks so that they could achieve more competitive results. After that, Huang et al. introduced the dense block (Fig. 1.(b)). Residual block and dense block use a single size of convolutional kernel and the computational complexity of dense blocks increases at a higher growth rate. In order to solve these drawbacks, we propose a multi-scale residual block.

Based on the residual structure, we introduce convolution kernels of diﬀerent sizes, which designed for adaptively detecting the features of images at diﬀerent scales. Meanwhile, a skip connection is applied between diﬀerent scale features so that the features information can be shared and reused with each other. This helps to fully exploit the local features of the image. In addition, a 1×1 convolution layer at the end of the block can be used as a bottleneck layer, which contributes to feature fusion and reduces computation complexity. We will give a more detailed description in section 3.1

**References**

[1] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deeply improved sparse coding for image super-resolution," *arXiv preprint arXiv:1507.08905,* vol. 2, no. 3, p. 4, 2015.

[2] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.