Mixed Link Networks

Wenhai Wang[1], Xiang Li[2], Jian Yang[2], Tong Lu[1]

[1] National Key Lab for Novel Software Technology, Nanjing University

[2] DeepInsight@PCALab, Nanjing University of Science and Technology

Abstract

On the basis of the analysis by revealing the equivalence of modern networks, we find that both ResNet and DenseNet are essentially derived from the same "dense topology", yet they only differ in the form of connection —— addition (dubbed "inner link") *vs.* concatenation (dubbed "outer link"). However, both forms of connections have the superiority and insufficiency. To combine their advantages and avoid certain limitations on representation learning, we present a highly efficient and modularized Mixed Link Network (MixNet) which is equipped with flexible inner link and outer link modules. Furthermore, we demonstrate that MixNets can achieve superior efficiency in parameter over the state-of-the-art architectures on many competitive datasets like CIFAR-10/100, SVHN and ImageNet.

Dense Topology

Let us consider a network that comprises L layers, each of which implements a non-linear transformation $H_\ell(\cdot)$, where ℓ indexes the layer. $H_\ell(\cdot)$ could be a composite function of several operations such as linear transformation, convolution, activation function, pooling, batch normalization. As illustrated in Fig. 1 (a), X_ℓ refers to the output of the transformation $H_\ell(\cdot)$ and S_ℓ is the result of the connection function $C(\cdot)$ whose inputs come from all the previous feature-maps X (i.e., $X_0, X_1, \dots, X_\ell$). Dense topology is defined as a path topology where each layer is connected with all the previous layers. Therefore, we can formulate the general form of dense topology simply as:

$$X_\ell = H_\ell\big(C(X_0, X_1, \cdots, X_{\ell-1})\big). \tag{1}$$

As shown in Fig. 1 (b)(c), both ResNet and DenseNet are essentially derived from dense topology. Fig. 1 (d) shows the path topology of the proposed MixNet.

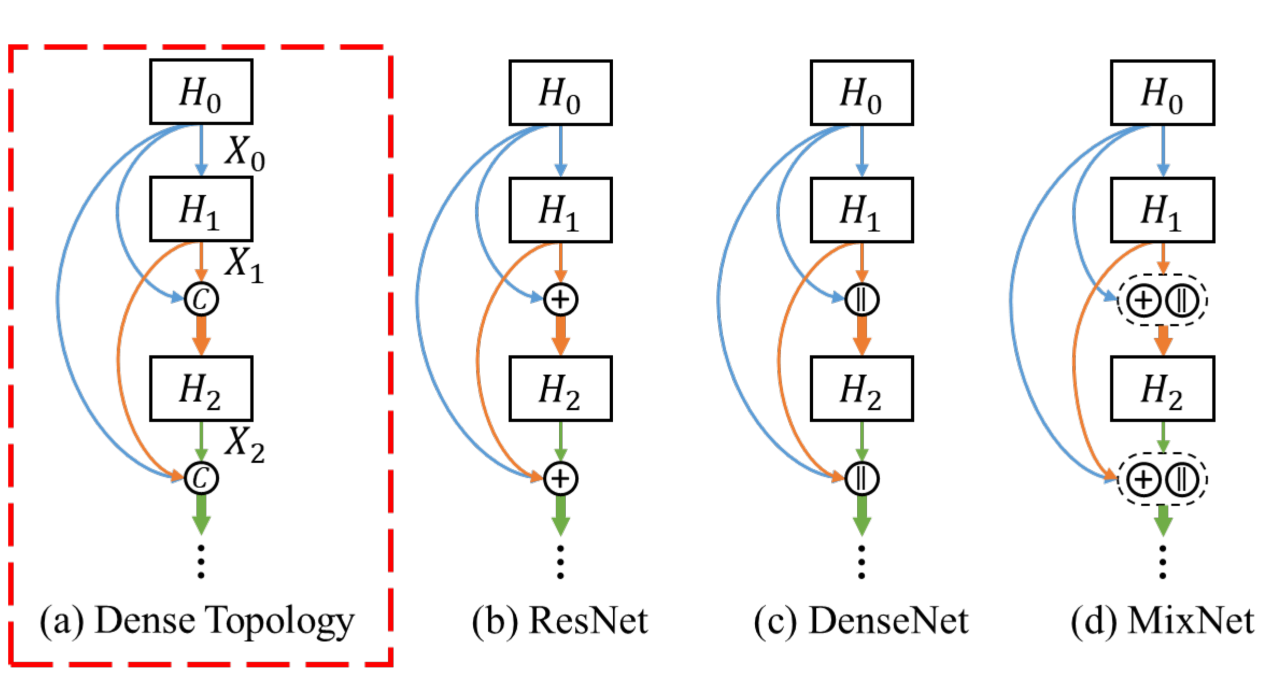

Fig. 1: The topological relations of different types of neural networks. The symbols "+" and "‖" denote element-wise addition and concatenation, respectively.

Mixed Link Networks

Inner/Outer Link Module

Let us define the output of Inner and Outer Link Module in ℓ-th layer as S_ℓ^{in} and S_ℓ^{out}, respectively. They can be formulated as:

$$S_\ell^{in} = S_{\ell-1}^{in} + H_\ell^{in}\big(S_{\ell-1}^{in}\big), \tag{2}$$

$$S_\ell^{out} = S_{\ell-1}^{out} \parallel H_\ell^{out}\big(S_{\ell-1}^{out}\big). \tag{3}$$

Mixed Link Architecture

We propose the mixed link architecture (see Fig.2) which embraces both inner link module and outer link module. The mixed link architecture can be formulated as Eqn. 4, a flexible combination of Eqn. 2 and Eqn. 3, to get a blending feature output S_ℓ:

$$S_\ell = \Big(S_{\ell-1} + H_\ell^{in}(S_{\ell-1})\Big) \parallel H_\ell^{out}(S_{\ell-1}). \tag{4}$$

It can be seen from Fig. 3 that the mixed link architecture with different configurations can reach four representative architectures (i.e. ResNet, DenseNet, DPN and our proposed MixNet).

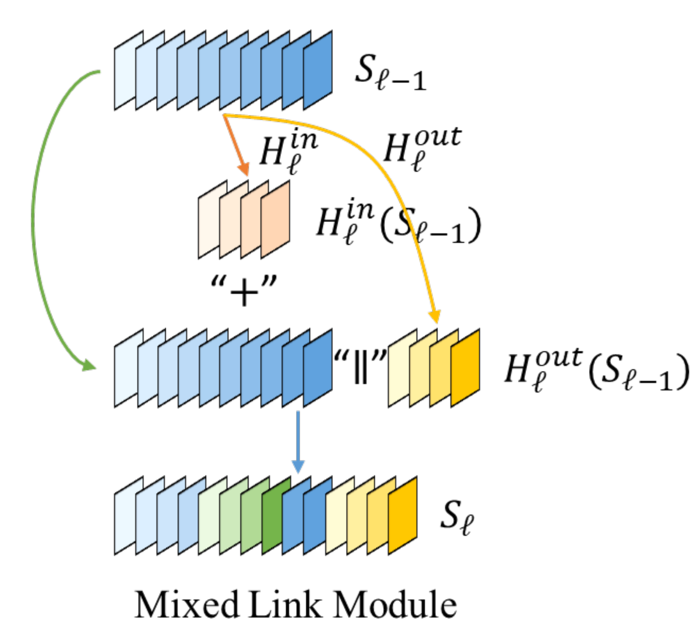

Fig. 2: The example of mixed link architecture. The symbol "+" and "‖" represent addition and concatenation.

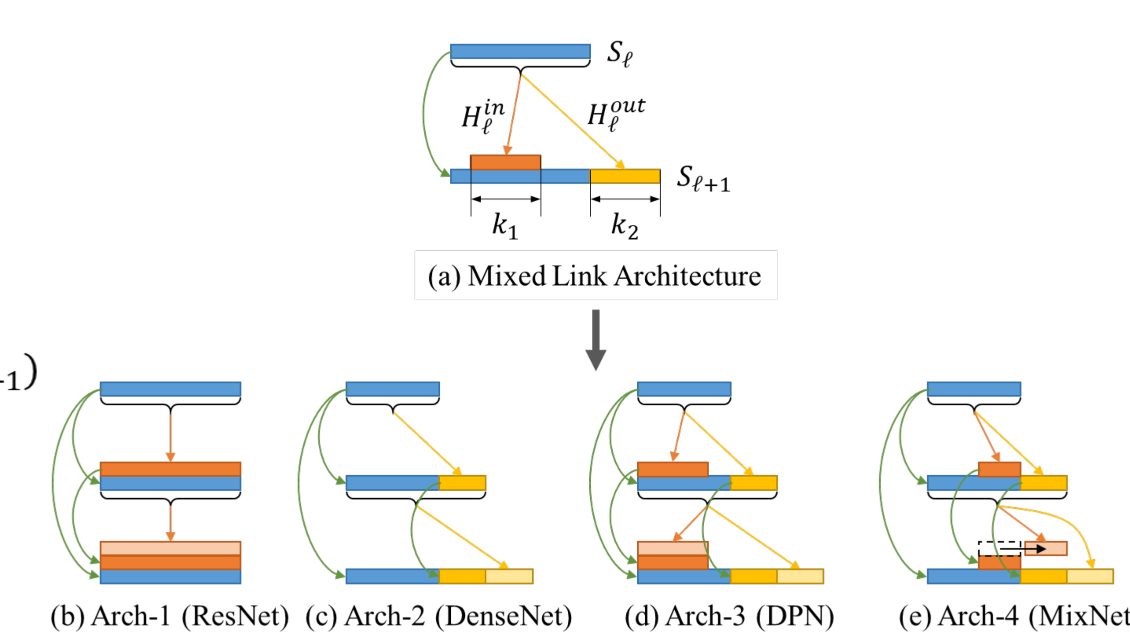

Fig. 3: Four architectures derived from mixed link architecture. The vertically aligned features are merged by element-wise addition, and the horizontally aligned features are merged by concatenation.

Experiment

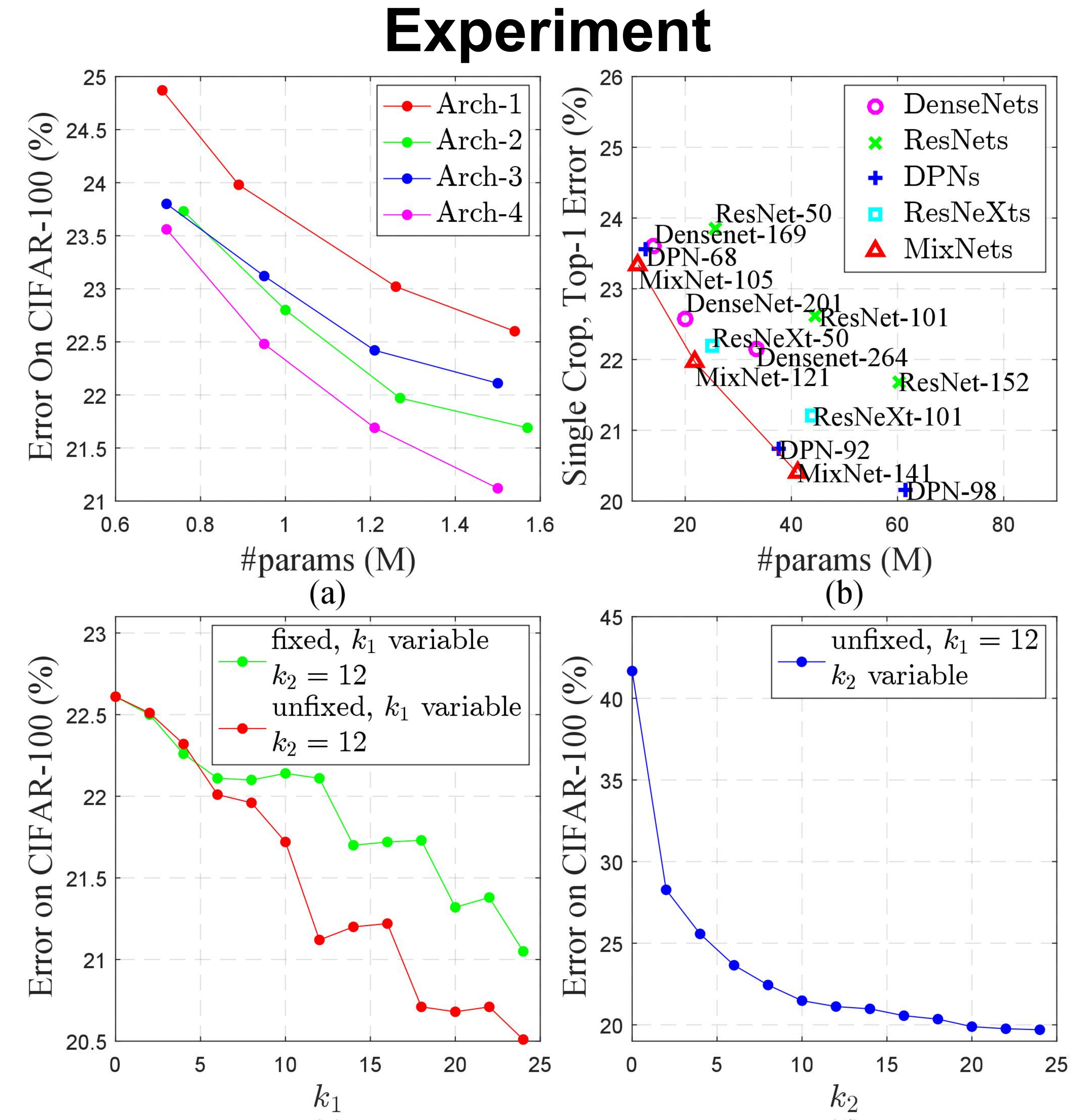

Fig.4: The illustrations of the experimental results. (a) Efficiency comparisons among the four architectures. (b) Comparisons comparison of the MixNets and the state-of-the-art architectures on ImageNet. (c) Fixed inner link *vs.* unfixed inner link. (d) The impact of outer link size.