

**1. How many layers are there in MLP Mixer? What are they and what is the purpose of these layers**

Two types of layers. Channel-mixing MLPs and Token-mixing MLPs. The channel-mixing MLPs allow communication between different channels and they operate on each token independently and take individual rows of the table as inputs. The token-mixing MLPs allow communication between different spatial locations (tokens) and they operate on each channel independently and take individual columns of the table as inputs.

**2. What are modern regularization schemes? Have authors used it in the paper? If yes, how and when?**

It is a key tool in the solution of inverse problems and it is used to introduce prior knowledge and allow a robust approximation of ill-posed pseudo inverses. Yes, the authors have used regularization techniques when pre-trained on ImageNet-21k that allowed the Mixer to achieve a strong overall performance otherwise the Mixer overfits without it. Mixer uses skip-connections and normalization layers.

**3. Which state art of the approaches are used to compare the result of experiments and what is the evaluation score?**

Convolutional Neural Networks and Vision Transformers. 87.94% top-1 validation accuracy on ILSVRC2012 "ImageNet".

**4. What methods have they used so their results outperformed the existing approaches ? Hint (Preprocessing , mathematical formulas , optimization )**

MLP layers, skip-connections, layer normalization, standard classification head with global average pooling layer followed by a linear classifier. Downstream tasks, pre-training, fine tuning.

**5. What sort of dataset is used by the authors? Please list the datasets they have used for their experiments?**

A large-scale hierarchical image database.

ILSVRC2021 ImageNet, and ImageNet-21k, a superset of ILSVRC2012 that contains 21k classes and 14M images . To assess performance at a larger scale, the authors train on JFT-300M, a proprietary dataset with 300M examples and 18k classes.

