# MLP-Mixer: An all-MLP Architecture for Vision (2021)

No CNNs, No Vision Transformers with Attention mechanism.

### Architecture

Per-batch fully connected, Mixer Layer and a classification layer.

### Methods:

1. Divide image into patches.
2. Fully connected layer to get latent representations.
3. Mixer Layer:
   1. Every patch has one vector.
   2. Layer Norm.
   3. Transpose channels and patches (channels, patches) instead of (patches, channels).
   4. MLP1: connecting the channels together from patches, and 2xfc layers + non-linear function.
   5. Transpose back channels and patches to MLP2.
   6. Use skip connections.
4. Global Average Pooling
5. Fully connected
6. Classification.

### Pros:

* Do not suffer from quadratic computational.
* Competitive but not state-of-the-art.
* Scales very well with training size.
* Very similar results with Transfer (ViT).
* Better than ResNet BiT.

# Involution: Inverting the Inherence of Convolution for Visual Recognition (2021)

Mix between CNN and Attention mechanism:

* Spatial-specific and channel agnostic.
* Shares kernel across all the channels.