**Complex Fusion Integration Module (CFIM)**

Illustrated in FigXX, the Complex Feature Interaction Module (CFIM) primarily encompasses two units connected in parallel. The Complex GlobalScope Multi-Layer Perceptron unit (CGMLP) is designed to capture and integrate the GlobalScope long-range dependencies of the real and imaginary parts in the uppermost feature XX. Simultaneously, for the aggregation of local area features within the real and imaginary layers, the complex local attention unit, deployed on XX, is introduced. The result feature maps from these two units are concatenate together along the channel dimension , forming CFIM's output for subsequent feature extraction. Between XX and the CFIM, a Stem block is utilized for feature smoothing, comprising two 5×5 convolution with 90 output channels for the real and imaginary part, followed by batch normalization layers and activation function layers. This process is encapsulated as follows:

Where X denotes the output of CFIM, and represent the output features of the Complex GlobalScope Multi-Layer Perceptron and the Complex Local Attention mechanism, respectively. is the output of the Stem block, which is obtained by

Where and indicate a 5×5 convolution operation with stride 1 and the channel size is configured to 90 for both the real and imaginary parts. represents a batch normalization layer and denotes the ReLU activation function. Finally , the real and imaginary results are concatenated along the last dimension.

1. **The Complex GlobalScope Multi-Layer Perceptron Unit(CGMLP)**

The CGMLP we proposed, mainly consisting of a depthwise convolution residual module, a channel MLP residual module and a Context Broadcasting module(CB).

**Depthwise Convolution Residual Module.** Post-smoothing in the Stem block, the features output is directed into depthwise convolution for both real and imaginary parts, followed by group normalizations (i.e., the real and imaginary feature maps are grouped along the channel dimension). Contrasted with the conventional spatial convolution for s Spectral feature extraction, the depthwise convolution enhances the spectral feature representation ability while reducing the computational costs. Subsequently, channel scaling operation is employed to obolster feature generalization and robustness[[1]](#endnote-1).The above processes, performed independently in both the real and imaginary parts, can be formulated as:

Where represents the output of Depthwise Convolution Residual Module. is the group normalization and is a depthwise convolution with the kernel size of 1×1.The denotes the channel scaling operation.

**Feature Interaction Module.** Following the channel scaling operation, we concatenate the real features maps and the imaginary features maps along the channel dimension. Subsequently, a 1×1 convolution is employed to facilitate the learning of compressed latent space representations of parameter distributions, enhancing cross-channel interaction[[2]](#endnote-2)[[3]](#endnote-3). These processes can be formulated as:

Where denotes a concatenation along the channel dimension, and denotes a 1×1 convolution.

**Channel MLP Residual Module.** Features from the Feature Interaction Unit, denoted as , are initially subjected to group normalization, followed by the application of the Channel Multi-Layer Perceptron (MLP)[[4]](#endnote-4). The Channel MLP can effectively learn the complex features of data through its multi-layer structure, concurrently reducing computational complexity[[5]](#endnote-5)[[6]](#endnote-6). Following this, a channel scaling operation is executed to enhance the generalization and robustness of features. The above processes are expressed as:

Where represents the output of Channel MLP Residual Module. is the group normalization and is the channel MLP. The denotes the channel scaling operation, and denotes the output of Feature Interaction Module.

**Context Broadcasting Module (CB).**

FigX illustrates our Context Broadcasting Module (CB) module. The CB module is positioned at the end of the Complex Long-range MLP block. See FigX for the overall CFIM architectures with our CB module. The CFIM is inclined towards dense interactions encompassing both real and imaginary parts, factoring in both long-range and short-range dependencies. However, due to the steep gradient of the softmax function, learning dense attention becomes more challenging. In contrast, the integration of the CB module mitigates the density in the original attention graph, thereby enhancing the capacity and generalization potential of the CFIM. Specifically, given a input ,our CB module supplies the average-pooled operation onto the channel features as follows:

Where is output of the Complex Long-range MLP block.denotes a average pooling operation on channel dimensions.

#### the Complex Local Attention Unit(CLA)

the Complex Local Attention Unit (CLA) operates as an encoder with an inherent dictionary mechanism. This encoder is composed of two primary elements:

The input feature initially undergoes processing in a combinatorial module, which encompasses two 1x1 convolutions and 3x3 convolutions. This module serves multiple purposes: compression of features, capturing spatial relationships, and expansion of channel information. The resultant transformed features are further refined through a CBR (Convolution-Batch Normalization-ReLU) module, which integrates a 3x3 convolutions, batch normalization, and ReLU activation. This integration is pivotal in enhancing feature characterization while preserving nonlinearity. Subsequently, the encoded features, denoted as are fed into the local attention unit. A set of scaling factor is utilized to align and map the corresponding position information. The comprehensive information of the entire features in relation to the k-th codeword is computed as follows:

Where is the update to the kth codebook vector. is a set of scaling factor. and are the real and imaginary parts of the i-th input vector, respectively. and are the real and imaginary parts of the k-th codebook vector, respectively. We use features fuse unit to fuse all and highlight key classes, which encompasses a batch Normalization (BN) layer , ReLU activation function , a mean layer and a fully connection layer. Building upon this, the full information of the whole *K* codewords is calculated as follows.

Where denotes BN layer , ReLU activation function and mean layer, and represents the fully connection layer.

**Feature Interaction Unit.** Upon acquiring the output from the features fuse unit, we introduce a feature interaction unit designed to facilitate the interaction between real and imaginary local features. Subsequently, a channel-wise multiplication is executed between the input features from the Stem block, denoted as and the scaling factor coefficient δ (·). Finally, we perform a channel-wise addition between features X\_smooth from the Stem block and the local attention features.The above processes are expressed as:

Where denotes the sigmoid function. is channel-wise multiplication and is the channel-wise addition.

1. W. Yu, M. Luo, P. Zhou, C. Si, Y. Zhou, X. Wang, J. Feng, and S. Yan,“Metaformer is actually what you need for vision,” arXiv, 2021. [↑](#endnote-ref-1)
2. Maksoud S, Zhao K, Peng C, et al. Scalable Bayesian Deep Learning with Kernel Seed Networks[J]. arXiv preprint arXiv:2104.09005, 2021. [↑](#endnote-ref-2)
3. Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo and Q. Hu, "ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks," 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 11531-11539, doi: 10.1109/CVPR42600.2020.01155. [↑](#endnote-ref-3)
4. Tolstikhin I O, Houlsby N, Kolesnikov A, et al. Mlp-mixer: An all-mlp architecture for vision[J]. Advances in neural information processing systems, 2021, 34: 24261-24272. [↑](#endnote-ref-4)
5. L. Nosrati, M. S. Fazel and M. Ghavami, "Improving Indoor Localization Using Mobile UWB Sensor and Deep Neural Networks," in IEEE Access, vol. 10, pp. 20420-20431, 2022, doi: 10.1109/ACCESS.2022.3151436. [↑](#endnote-ref-5)
6. Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, “Yolox: Exceeding yolo series

   in 2021,” arXiv, 2021. [↑](#endnote-ref-6)