Bing

Fusing is adding the skip connection with the recalibrated tensor, which is the result of applying the attention to the input tensor. The skip connection helps to preserve the original information of the input tensor, and to avoid losing too much detail or diversity in the output tensor.

Sparsity and singularity are two problems that can affect the quality and efficiency of neural networks. They are caused by the attention alone because the attention reduces the number of channels and the spatial dimensions of the input tensor, and also applies sigmoid functions that can make the values very close to zero or one.

Sparsity means that many values in the tensor are zero or close to zero. This can make the tensor less informative and less useful for the network. For example, if a tensor has only zeros, it means that the network does not pay any attention to the input at all, and it cannot learn anything from it.

Singularity means that the tensor becomes very sensitive to small changes, and it cannot be inverted or solved easily. This can make the network unstable and hard to train. For example, if a tensor has only ones, it means that the network pays too much attention to the input, and it cannot distinguish between different features or inputs.

You can find more information about sparsity and singularity in some web search results . I hope this helps you understand sparsity and singularity better. 😊