A Overall Process Pseudocode

As follows, is the pseudo code of the method in this paper.

Algorithm 1: Employ dynamic neural networks to compress images of different complexities with implicit neural representations

Data: Set of images
Result: M_ψ and Base model f_θ

image set's $G_{mean} \leftarrow$ Sobel operator ;
image complexity classification $\leftarrow$ K-means[G_{mean}] ;
initialization Coordinates x and $M_\psi = (Q_{d\times r}, Q^i_{r\times r}, V^T_{d\times r}), i = 1, \ldots, l$ approximate $r \propto G_{mean}$, $l \propto G_{mean}$;
while *Training:Taking a complexity level t image y_i set as an example.* **do**
    $z^0 \leftarrow \gamma(x)$;
    for *Traverse through the layers 1 to l in the base model f_θ* **do**
        $M^i_{soft} \leftarrow sigmoid(Q_{d\times r} \times Q^i_{r\times r} \times V^T_{d\times r})$
        $M_i \leftarrow \begin{cases} 1, M^i_{soft} \geq 0.5 \\ 0, M^i_{soft} < 0.5 \end{cases}$
        $z^i \leftarrow \sin(\omega_0((W^i \odot M^i)z^{i-1} + b^i))$
    end
    $f_\theta\,(x, M_\psi, t) \leftarrow W^L_t z^{L-1} + b^L$
    $L_{mse} \leftarrow ||f_\theta(x, M_\psi, t), y_i||_2$
    update the parameters using like-meta-learning
end
quantize M_ψ and the base model f_θ

B Class Meta-learning M_ψ

Store each image in the dataset explicitly for its M_ψ and evaluate each coordinate point when reconstructing the image. Denote the output y as

$$y = f_\theta\,(x, M_\psi, t) \tag{8}$$

Only M_ψ of each image alone needs to be overfitted to each image by the base network. Also, minimize the difference between the reconstruction result of the base dictionary model on the whole dataset and the dataset data.

$$L(x, M_\psi, d) = \sum_{j=1}^{n} ||f_\theta(x_i, M_\psi, t), y_i||_2 \tag{9}$$

$$\min_{\theta, M_\psi} L(x, M_\psi, d) \tag{10}$$

COIN++[7], a special MAML is applied to generate well-initialized network parameters. while COIN++ emphasizes the generalization of the model to obtain updated network parameters by several gradient descent, in our experiments, multi-task learning is performed to overfit M_ψ. It is required to meta-learning a θ that over-fits the storage matrix M_ψ at each new data point. Therefore, in the inner loop, M_ψ is learned in the following way:

$$M_\psi^{(j)} = M_\psi^{(j)} - \alpha \nabla_{M_\psi} L\left(\theta, M_\psi, d^{(j)}\right) \tag{11}$$

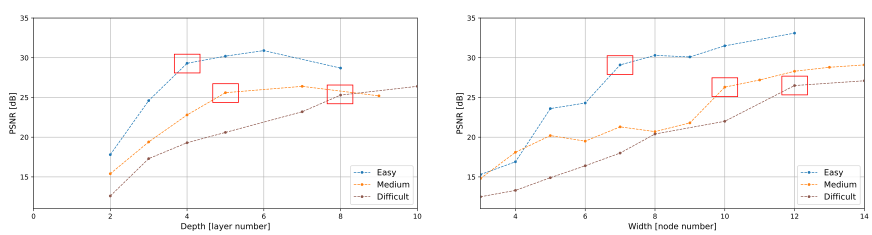
In the outer loop, the network parameters θ are updated using the errors generated for each data point:

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_{j=1}^{N} L\left(\theta, M_\psi^{(j)}, d^{(j)}\right) \tag{12}$$

The outer loop base model learns as many dictionary frequencies as possible, while the inner loop performs gradient updates for each image's storage matrix, guiding the generation mask to correctly select the appropriate frequency in the base dictionary to achieve multi-image compression.

C Experimental details

C.1 Verify the Need For Dynamic Networks



(a) the necessity of dynamic depth of neural network

(b) the necessity of dynamic width of neural network

Fig. 4. Verify the dynamic necessity of neural network. Take twenty pictures of each complexity and change the depth or width to get the image reconstruction effect of each complexity picture in a specific network, which is measured by PSNR.

C.2 Analysis of Experimental Results of Ablation Dynamic Depth

We selected 20 images from each complexity category, maintaining their original width (r) within the category. We then calculated the average image reconstruction effects while increasing the number of network layers.

According to the figure5(a), as the complexity of the images gradually increased, we observed that a greater number of network layers were required to

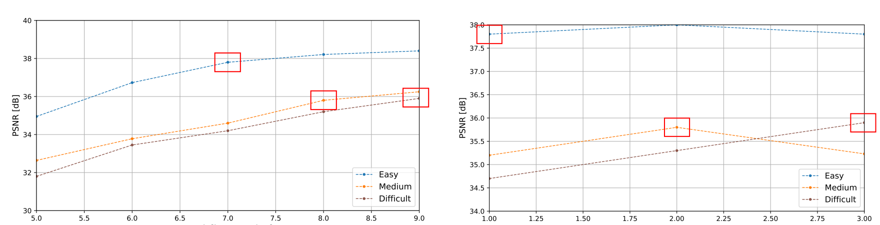
Table 2. Fixed width and dynamic depth image reconstruction effect table with different complexity

Network Layers	Simple Image	Middle Image	Complex Image
5	34.95	32.64	31.8
6	36.73	33.78	33.45
7	37.8	34.6	34.2
8	38.21	35.8	35.2
9	38.4	36.25	35.9

achieve a better image reconstruction effect while consuming relatively more resources. For instance, easy images yielded the best effect at 7 layers, medium images at 8 layers, and difficult images required a deeper network with the best effect achieved at 9 layers. This demonstrates that the appropriate network depth varies depending on the complexity of the images. Deeper networks are necessary for more complex images to achieve superior image reconstruction. This further affirms the importance of dynamically adapting the network depth in our approach, allowing for optimal image reconstruction results across images of different complexities.

Our analysis confirms the significance of dynamically adjusting the network depth in our approach, achieving optimal image reconstruction results for images of different complexities.

C.3 Analysis of Experimental Results of Ablation Dynamic Width



(a) Ablation dynamic depth (b) Ablation dynamic width

Fig. 5. The results of ablation experiments on dynamic neural networks. During the ablation of dynamic depth, the fixed network width, that is, the fixed R-value, gradually deepens the network depth. Compare the optimal network depth for images of different complexity. When the dynamic width is ablated, the network depth is fixed and the network width is gradually widened, that is, the R-value is increased. Compare the optimal network width for images of different complexity.

For each category, we selected 20 images of each and fixed their optimal depths (verified in previous experiments). Their average image reconstruction effect values were calculated with increasing network width (by adjusting the r value).

According to the figure5(b), we can intuitively observe that the best results are achieved with an r value of 2 in medium complexity images and with an r value of 3 in difficult complexity images, both with the same width setting. In the case of simple complexity images, as the r value increases, i.e., the network width becomes wider, the effect of image reconstruction increases slightly and then decreases to the same as the initial one. This is because the simple image requires less frequency composition, while the increase of network width brings more computational consumption, and the optimization effect is not obvious. Therefore the optimal r value for the simple image should be 1, again with the same setting as its width.

Specifically, our set the following r values: r=1 (easy image: 37.8, medium image: 35.2, difficult image: 37.8), r=2 (easy image: 38.03, medium image: 35.88, difficult image: 35.23), r=3 (easy image: 37.8, medium image: 35.3, difficult image: 35.9). From the image reconstruction results with different r values, it can be intuitively seen that the best results are achieved with a r value of 2 in the medium images and with a r value of 3 in the difficult images, which coincides with their width settings. In contrast, in the easy images, as the r value increases (i.e., the network width increases), the image reconstruction effect first slightly improves and then decreases to the same as the initial one. This is because the easy image requires less frequency composition [33], while increasing the network width at the same time adds more computational overhead and is less effective in optimization. Therefore, the optimal r value for simple images should be 1, which is the same as its initial width. And the optimal r values for medium and difficult images are 2 and 3, respectively, and their image reconstruction effects under other r values follow the variation trend.

These results analyzed further validate the necessity of dynamically adapting the network width in our method and being able to adapt to images of different complexity to achieve the best image reconstruction results.