当然可以！以下是我为你整理的一份高质量的 Deep Learning General Questions，风格与你考试那张非常接近，但范围更全面，涵盖了优化器、初始化、正则化、梯度、激活函数、损失函数、训练技巧等多个角度，每题都可以简答（1-2 句话）。适合作为查漏补缺的练习。

**🧠 Deep Learning General Questions（查漏补缺题）**

**🔧 Optimization / Training**

1. What is the main purpose of using momentum in gradient descent?
2. How does Adam optimizer differ from SGD?
3. Why is learning rate scheduling important during training?
4. What happens if the learning rate is too small?
5. What does gradient clipping help prevent?

**📉 Loss Functions**

1. Which loss function is commonly used in binary classification tasks?
2. What is the difference between MSE and MAE in regression tasks?
3. Why is cross-entropy preferred over MSE for classification?

**🔍 Initialization**

1. Why is proper weight initialization important in deep networks?
2. What initialization method is commonly used with ReLU activations?
3. What problem can arise if all weights are initialized to zero?

**🎛️ Activation Functions**

1. Why is ReLU preferred over sigmoid in deep networks?
2. What is the dying ReLU problem?
3. What is the benefit of using Leaky ReLU?

**🎯 Regularization / Overfitting**

1. What does L2 regularization do to the weights?
2. How does dropout help prevent overfitting?
3. What is early stopping and how does it prevent overfitting?

**⛓️ Backpropagation / Gradients**

1. What is vanishing gradient problem and where does it typically occur?
2. Which activation functions are more prone to vanishing gradients?
3. How does batch normalization help with gradient flow?

**📦 Batch / Data**

1. What is the role of a mini-batch in training neural networks?
2. How can batch size affect generalization?
3. Why do we shuffle the data before each epoch?

**🧩 Model Architecture / Training Tricks**

1. What is residual connection and why is it used?
2. What problem do skip connections help solve in deep networks?
3. What is layer normalization and when is it preferred over batch norm?
4. Why do we use weight decay?

**💻 Computational Considerations**

1. Why is GPU often used for deep learning training?
2. What does it mean if training loss decreases but validation loss increases?

**📚 Advanced / Bonus**

1. What is label smoothing and why is it used?
2. In what scenario is transfer learning particularly useful?
3. What are exploding gradients and how can they be handled?