Problem 1

Visualize the similarity between different pairs of positional embedding and briefly explain the result.

```
pos_emb = model.decoder.embed_positions.weights.cpu().detach()
print(pos_emb.shape)
```

[18] ✓ 0.5s Python

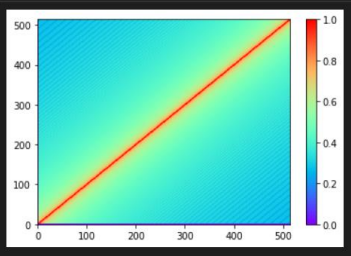
```
torch.Size([514, 512])
```

```
cm=plt.cm.get_cmap('rainbow')
z = []
for x1 in pos_emb:
    for x2 in pos_emb:
        z.append(torch.nn.functional.cosine_similarity(x1, x2, dim=0, eps=1e-08))
```

[40] ✓ 3.9s Python

```
plt.pcolor(np.array(z).reshape(514,514),cmap=cm)
plt.colorbar()
plt.show()
```

[45] ✓ 1.9s Python



Positional encoding 的用途是 encode position 的資訊，所以 position 比較近的 embedding 的 similarity 會比較高，反之比較遠則比較小，所以可以看到圖中對角線的 similarity 為 1。

Problem 2

Clip gradient norm and visualize the changes of gradient norm in different steps.

Circle two places with gradient explosion.

```
data = [x.item() for x in gnorm_record]
```

✓ 0.9s

```
plt.plot(data)
plt.show()
```

✓ 0.1s