

原始

```
Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=2.0, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.2, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```

官方给的模型：

```
Loading dataset
Computing results...
Test: [0/40]    Le 65.7924 (65.7924)    Time 0.678 (0.000)
Test: [10/40]   Le 65.6336 (66.4297)    Time 0.019 (0.000)
Test: [20/40]   Le 65.4652 (66.5281)    Time 0.018 (0.000)
Test: [30/40]   Le 68.9066 (67.2639)    Time 0.022 (0.000)
Images: 1000, Captions: 5000
rsum: 272.8
Average i2t Recall: 50.8
Image to text: 28.7 56.4 67.2 4.0 35.1
Average t2i Recall: 40.2
Text to image: 19.9 44.7 55.9 7.0 44.1
```

我自己的最好模型：

```
Loading dataset
Computing results...
Test: [0/40]    Le 69.5436 (69.5436)    Time 0.620 (0.000)
Test: [10/40]   Le 69.6387 (67.8007)    Time 0.021 (0.000)
Test: [20/40]   Le 65.7297 (67.6416)    Time 0.020 (0.000)
Test: [30/40]   Le 67.0183 (68.2584)    Time 0.024 (0.000)
Images: 1000, Captions: 5000
rsum: 280.7
Average i2t Recall: 51.2
Image to text: 29.7 55.7 68.3 4.0 30.2
Average t2i Recall: 42.3
Text to image: 21.9 47.2 57.9 6.0 39.4
```

微调

改变cnn_type

cnn type=resnet152

```
Namespace(batch_size=128, cnn_type='resnet152', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=2.0, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.2, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```

```

Loading dataset
Computing results...
Test: [0/40]    Le 68.0888 (68.0887)    Time 0.635 (0.000)
Test: [10/40]   Le 69.0510 (67.3034)    Time 0.019 (0.000)
Test: [20/40]   Le 67.2831 (67.2574)    Time 0.018 (0.000)
Test: [30/40]   Le 68.6796 (68.0767)    Time 0.021 (0.000)
Images: 1000, Captions: 5000
rsum: 278.9
Average i2t Recall: 50.6
Image to text: 29.4 56.9 65.5 4.0 27.7
Average t2i Recall: 42.4
Text to image: 21.7 46.7 58.7 7.0 38.9

```

改变margin

margin=0.1

```

Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=
2.0, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.1, max_violation=True, measure='cosine',
no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/'
, word_dim=300, workers=10)

```

```

Loading dataset
Computing results...
Test: [0/40]    Le 37.0786 (37.0785)    Time 0.646 (0.000)
Test: [10/40]   Le 35.0437 (35.6524)    Time 0.020 (0.000)
Test: [20/40]   Le 34.9152 (35.7826)    Time 0.018 (0.000)
Test: [30/40]   Le 34.5568 (35.8964)    Time 0.022 (0.000)
Images: 1000, Captions: 5000
rsum: 280.3
Average i2t Recall: 51.5
Image to text: 29.0 58.2 67.4 4.0 29.2
Average t2i Recall: 41.9
Text to image: 21.0 46.8 57.9 6.0 38.2

```

margin=0.3

```

Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=
2.0, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.3, max_violation=True, measure='cosine',
no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/'
, word_dim=300, workers=10)

```

```

Loading dataset
Computing results...
Test: [0/40]    Le 95.2547 (95.2546)    Time 0.615 (0.000)
Test: [10/40]   Le 94.9247 (94.9285)    Time 0.020 (0.000)
Test: [20/40]   Le 92.7183 (94.3758)    Time 0.018 (0.000)
Test: [30/40]   Le 96.8862 (95.4214)    Time 0.020 (0.000)
Images: 1000, Captions: 5000
rsum: 284.4
Average i2t Recall: 52.1
Image to text: 30.0 57.6 68.7 4.0 29.8
Average t2i Recall: 42.7
Text to image: 21.7 47.2 59.2 6.0 41.0

```

margin=0.4

```
Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=2.0, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.4, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```

```
Loading dataset
Computing results...
Test: [0/40]    Le 121.2568 (121.2567)  Time 0.634 (0.000)
Test: [10/40]   Le 120.3909 (119.9962)  Time 0.020 (0.000)
Test: [20/40]   Le 119.6137 (119.9944)  Time 0.018 (0.000)
Test: [30/40]   Le 119.3687 (120.7050)  Time 0.022 (0.000)
Images: 1000, Captions: 5000
rsum: 287.1
Average i2t Recall: 52.6
Image to text: 31.9 57.5 68.3 4.0 29.9
Average t2i Recall: 43.1
Text to image: 21.8 48.1 59.5 6.0 39.9
```

margin=0.5

```
Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=2.0, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.5, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```

```
Loading dataset
Computing results...
Test: [0/40]    Le 145.5935 (145.5934)  Time 0.658 (0.000)
Test: [10/40]   Le 147.3819 (145.8561)  Time 0.021 (0.000)
Test: [20/40]   Le 145.7023 (145.9441)  Time 0.019 (0.000)
Test: [30/40]   Le 147.7073 (146.6563)  Time 0.023 (0.000)
Images: 1000, Captions: 5000
rsum: 282.6
Average i2t Recall: 51.8
Image to text: 31.5 56.5 67.4 4.0 30.2
Average t2i Recall: 42.4
Text to image: 22.3 46.7 58.2 7.0 41.8
```

margin=0.6

```
Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=2.0, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.6, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```

```
Loading dataset
Computing results...
Test: [0/40]    Le 173.7031 (173.7029)  Time 0.623 (0.000)
Test: [10/40]   Le 174.9549 (172.5871)  Time 0.019 (0.000)
Test: [20/40]   Le 169.5366 (172.4697)  Time 0.019 (0.000)
Test: [30/40]   Le 174.6542 (173.4704)  Time 0.023 (0.000)
Images: 1000, Captions: 5000
rsum: 284.9
Average i2t Recall: 52.1
Image to text: 30.3 56.7 69.4 4.0 35.4
Average t2i Recall: 42.8
Text to image: 21.4 47.8 59.3 6.0 45.4
```

margin=1.0

```
Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=2.0, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=1.0, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```

```
Loading dataset
Computing results...
Test: [0/40]    Le 273.8957 (273.8954)   Time 0.637 (0.000)
Test: [10/40]   Le 273.7250 (273.1289)   Time 0.021 (0.000)
Test: [20/40]   Le 270.8572 (273.2609)   Time 0.019 (0.000)
Test: [30/40]   Le 273.5240 (273.9809)   Time 0.023 (0.000)
Images: 1000, Captions: 5000
rsum: 282.1
Average i2t Recall: 51.2
Image to text: 29.6 55.8 68.2 4.0 30.4
Average t2i Recall: 42.8
Text to image: 22.2 47.4 58.9 6.0 37.3
```

改变num_epochs与lr_update

num_epochs=60 lr_update=30

```
Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=2.0, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=30, margin=0.4, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=60, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```

```
Loading dataset
Computing results...
Test: [0/40]    Le 121.0823 (121.0822)   Time 0.628 (0.000)
Test: [10/40]   Le 120.3664 (120.2793)   Time 0.019 (0.000)
Test: [20/40]   Le 118.5475 (120.2287)   Time 0.018 (0.000)
Test: [30/40]   Le 121.8343 (121.1043)   Time 0.022 (0.000)
Images: 1000, Captions: 5000
rsum: 283.2
Average i2t Recall: 51.9
Image to text: 30.3 57.2 68.1 4.0 28.7
Average t2i Recall: 42.5
Text to image: 21.7 47.2 58.7 6.0 41.3
```

改变num_epochs

num_epochs=40

```
Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=2.5, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.4, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=40, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```

```

Loading dataset
Computing results...
Test: [0/40]    Le 120.3024 (120.3023)  Time 0.750 (0.000)
Test: [10/40]   Le 121.6175 (120.5140)  Time 0.026 (0.000)
Test: [20/40]   Le 119.9841 (120.5643)  Time 0.019 (0.000)
Test: [30/40]   Le 123.1514 (121.5071)  Time 0.023 (0.000)
Images: 1000, Captions: 5000
rsum: 284.2
Average i2t Recall: 52.0
Image to text: 29.9 57.1 69.1 4.0 31.9
Average t2i Recall: 42.7
Text to image: 22.2 47.2 58.6 6.0 44.3

```

改变grad_clip

grad_clip=1.5

```

Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=
1.5, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.4, max_violation=True, measure='cosine',
no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/'
, word_dim=300, workers=10)

```

```

Loading dataset
Computing results...
Test: [0/40]    Le 121.9593 (121.9592)  Time 0.673 (0.000)
Test: [10/40]   Le 120.1685 (120.3988)  Time 0.020 (0.000)
Test: [20/40]   Le 118.8767 (120.7569)  Time 0.020 (0.000)
Test: [30/40]   Le 122.0564 (121.5253)  Time 0.022 (0.000)
Images: 1000, Captions: 5000
rsum: 283.7
Average i2t Recall: 51.7
Image to text: 29.8 56.5 68.7 4.0 31.8
Average t2i Recall: 42.9
Text to image: 21.4 47.9 59.4 6.0 45.0

```

grad_clip=2.3

```

Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=
2.3, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.4, max_violation=True, measure='cosine',
no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/'
, word_dim=300, workers=10)

```

```

Loading dataset
Computing results...
Test: [0/40]    Le 121.8544 (121.8543)  Time 0.622 (0.000)
Test: [10/40]   Le 120.3670 (120.5463)  Time 0.021 (0.000)
Test: [20/40]   Le 120.6932 (120.7348)  Time 0.017 (0.000)
Test: [30/40]   Le 121.6854 (121.7728)  Time 0.011 (0.000)
Images: 1000, Captions: 5000
rsum: 286.3
Average i2t Recall: 52.3
Image to text: 30.2 57.8 68.9 4.0 33.6
Average t2i Recall: 43.1
Text to image: 22.4 48.0 59.1 6.0 43.9

```

grad_clip=2.4


```
Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=2.4, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.4, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```

```
Loading dataset
Computing results...
Test: [0/40]    Le 121.1220 (121.1219)  Time 0.748 (0.000)
Test: [10/40]   Le 120.3777 (120.5296)  Time 0.027 (0.000)
Test: [20/40]   Le 122.4092 (120.5811)  Time 0.024 (0.000)
Test: [30/40]   Le 121.6419 (121.2731)  Time 0.028 (0.000)
Images: 1000, Captions: 5000
rsum: 280.7
Average i2t Recall: 51.0
Image to text: 30.6 56.0 66.5 4.0 30.1
Average t2i Recall: 42.5
Text to image: 21.5 47.1 59.0 6.0 41.3
```

grad_clip=2.45

```
Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=2.45, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.4, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```

```
Loading dataset
Computing results...
Test: [0/40]    Le 119.3939 (119.3938)  Time 0.660 (0.000)
Test: [10/40]   Le 119.8524 (119.5639)  Time 0.019 (0.000)
Test: [20/40]   Le 118.9205 (119.3947)  Time 0.017 (0.000)
Test: [30/40]   Le 120.7806 (120.2720)  Time 0.021 (0.000)
Images: 1000, Captions: 5000
rsum: 283.2
Average i2t Recall: 51.7
Image to text: 29.7 57.7 67.7 4.0 28.5
Average t2i Recall: 42.7
Text to image: 22.1 47.2 58.8 7.0 38.4
```

grad_clip=2.5

```
Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=2.5, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.4, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```

```
Loading dataset
Computing results...
Test: [0/40]    Le 121.5946 (121.5945)  Time 0.674 (0.000)
Test: [10/40]   Le 120.7794 (120.8481)  Time 0.020 (0.000)
Test: [20/40]   Le 119.7017 (120.8967)  Time 0.021 (0.000)
Test: [30/40]   Le 122.1823 (121.8101)  Time 0.024 (0.000)
Images: 1000, Captions: 5000
rsum: 290.4
Average i2t Recall: 53.5
Image to text: 32.7 57.3 70.5 4.0 34.2
Average t2i Recall: 43.3
Text to image: 22.5 48.4 59.1 6.0 45.7
```

grad_clip=2.55

```
Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=2.55, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.4, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```

```
Loading dataset
Computing results...
Test: [0/40]    Le 121.6188 (121.6187)  Time 0.704 (0.000)
Test: [10/40]   Le 122.0213 (121.1525)  Time 0.020 (0.000)
Test: [20/40]   Le 120.6680 (121.0394)  Time 0.019 (0.000)
Test: [30/40]   Le 121.6718 (121.9225)  Time 0.023 (0.000)
Images: 1000, Captions: 5000
rsum: 283.2
Average i2t Recall: 51.4
Image to text: 30.8 55.9 67.5 4.0 34.6
Average t2i Recall: 43.0
Text to image: 22.1 47.7 59.2 6.0 46.1
```

grad_clip=2.6

```
Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=2.6, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.4, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```

```
Loading dataset
Computing results...
Test: [0/40]    Le 119.9080 (119.9079)  Time 0.630 (0.000)
Test: [10/40]   Le 119.3398 (119.2000)  Time 0.018 (0.000)
Test: [20/40]   Le 119.5898 (119.3093)  Time 0.018 (0.000)
Test: [30/40]   Le 121.7911 (119.9375)  Time 0.021 (0.000)
Images: 1000, Captions: 5000
rsum: 281.8
Average i2t Recall: 51.0
Image to text: 28.5 55.9 68.7 4.0 30.1
Average t2i Recall: 42.9
Text to image: 21.4 47.8 59.4 6.0 38.1
```

grad_clip=3.0

```
Namespace(batch_size=128, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=3.0, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.4, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```

```
Loading dataset
Computing results...
Test: [0/40]    Le 121.1723 (121.1722)  Time 0.634 (0.000)
Test: [10/40]   Le 120.4878 (119.7524)  Time 0.020 (0.000)
Test: [20/40]   Le 119.8523 (119.7264)  Time 0.019 (0.000)
Test: [30/40]   Le 119.6936 (120.3449)  Time 0.022 (0.000)
Images: 1000, Captions: 5000
rsum: 283.8
Average i2t Recall: 51.6
Image to text: 30.7 56.2 68.0 4.0 27.7
Average t2i Recall: 43.0
Text to image: 21.9 48.0 59.1 6.0 39.1
```

改batch_size

batch_size=256

```
Namespace(batch_size=256, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=2.5, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.4, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```

```
Loading dataset
Computing results...
Test: [0/20]    Le 239.8498 (239.8497)   Time 0.738 (0.000)
Test: [10/20]   Le 242.1881 (240.5225)   Time 0.040 (0.000)
Images: 1000, Captions: 5000
rsum: 289.3
Average i2t Recall: 53.3
Image to text: 31.8 59.4 68.8 3.0 29.5
Average t2i Recall: 43.1
Text to image: 22.6 47.1 59.5 6.0 39.9
```

batch_size=64

```
Namespace(batch_size=64, cnn_type='vgg19', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=2.5, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.4, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```

```
Computing results...
Test: [0/79]    Le 58.5200 (58.5199)     Time 0.624 (0.000)
Test: [10/79]   Le 60.4370 (59.5277)     Time 0.011 (0.000)
Test: [20/79]   Le 60.0492 (59.6634)     Time 0.013 (0.000)
Test: [30/79]   Le 59.1038 (59.7039)     Time 0.012 (0.000)
Test: [40/79]   Le 57.9783 (59.5058)     Time 0.012 (0.000)
Test: [50/79]   Le 60.0398 (59.6157)     Time 0.012 (0.000)
Test: [60/79]   Le 58.2084 (59.8674)     Time 0.015 (0.000)
Test: [70/79]   Le 59.5572 (59.9560)     Time 0.012 (0.000)
Images: 1000, Captions: 5000
rsum: 288.7
Average i2t Recall: 51.8
Image to text: 30.4 57.0 67.9 4.0 40.4
Average t2i Recall: 44.5
Text to image: 22.8 49.1 61.4 6.0 48.3
```

再改成resnet152

```
Namespace(batch_size=128, cnn_type='resnet152', crop_size=224, data_name='f30k_precomp', data_path='./data', embed_size=1024, finetune=False, grad_clip=2.5, img_dim=4096, learning_rate=0.0002, log_step=10, logger_name='runs/f30k_vse++', lr_update=15, margin=0.4, max_violation=True, measure='cosine', no_imgnorm=False, num_epochs=30, num_layers=1, reset_train=True, resume='', use_abs=False, use_restval=False, val_step=500, vocab_path='./vocab/', word_dim=300, workers=10)
```



```
Loading dataset
Computing results...
Test: [0/40]    Le 122.1266 (122.1265)  Time 0.612 (0.000)
Test: [10/40]   Le 120.9034 (120.8604)  Time 0.019 (0.000)
Test: [20/40]   Le 120.0100 (120.5763)  Time 0.018 (0.000)
Test: [30/40]   Le 122.4724 (121.3270)  Time 0.021 (0.000)
Images: 1000, Captions: 5000
rsum: 285.5
Average i2t Recall: 52.1
Image to text: 30.0 58.3 68.1 4.0 33.6
Average t2i Recall: 43.0
Text to image: 22.0 48.1 59.1 6.0 41.6
```