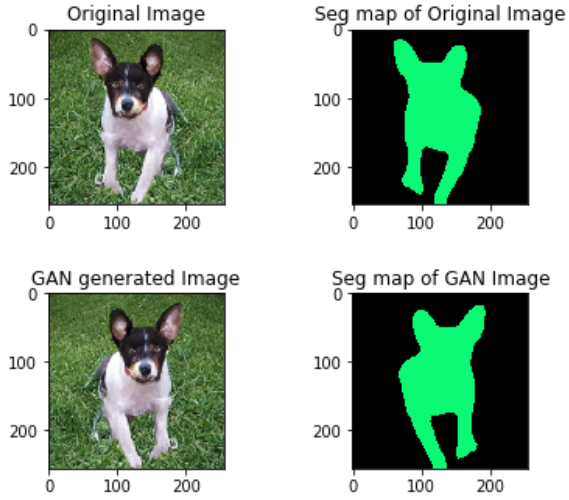
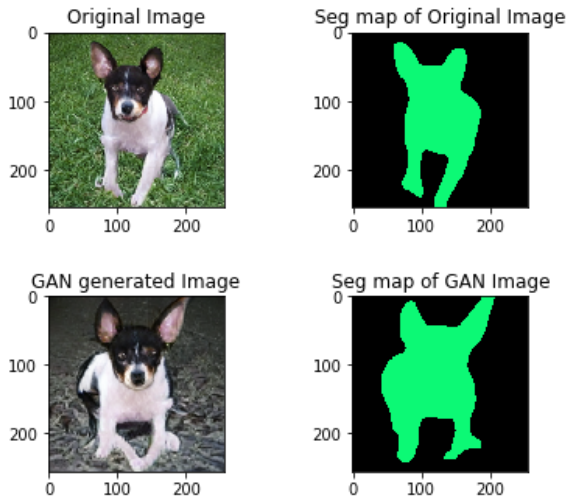
```
    print('-'*100)
    cost.backward()
    optimizer.step()

print('End')
```



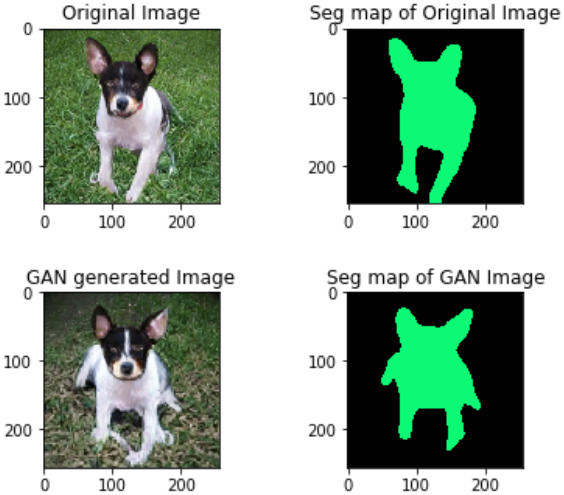
```
cost_old 1  is  tensor(1.1838, device='cuda:0', grad_fn=<SqrtBackw
ard>)
cost 1  is  tensor(0.6006, device='cuda:0', grad_fn=<NllLoss2DBack
ward>)
```



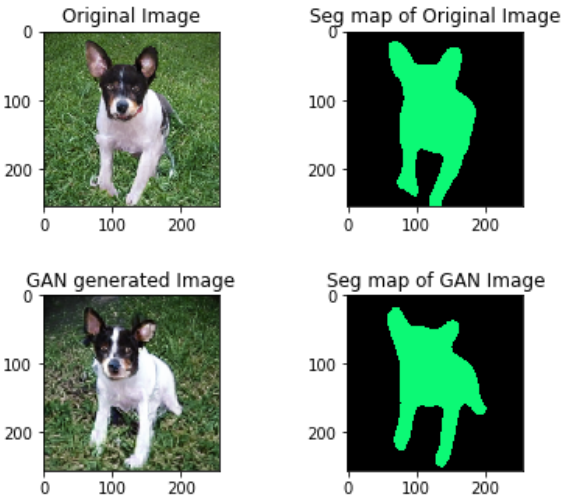

```
--------------------------------------------------------------------
---------------------------------
cost_old 2  is  tensor(1.6854, device='cuda:0', grad_fn=<SqrtBackw
ard>)
cost 2  is  tensor(0.4110, device='cuda:0', grad_fn=<NllLoss2DBack
ward>)
```



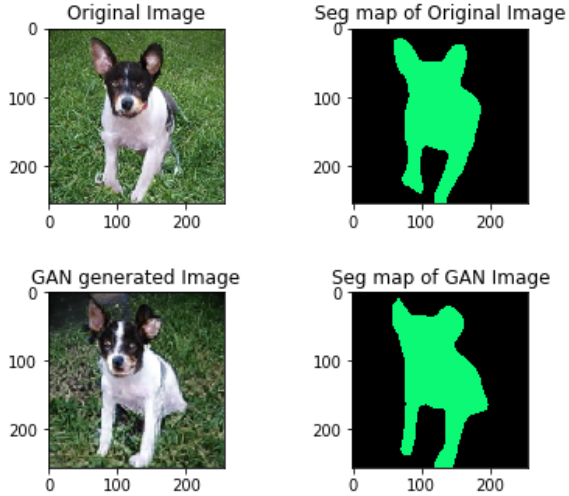
```
----------------------------------------------------------------------
----------------------------------
cost_old 3  is  tensor(1.5548, device='cuda:0', grad_fn=<SqrtBackw
ard>)
cost 3  is  tensor(0.2415, device='cuda:0', grad_fn=<NllLoss2DBack
ward>)
```



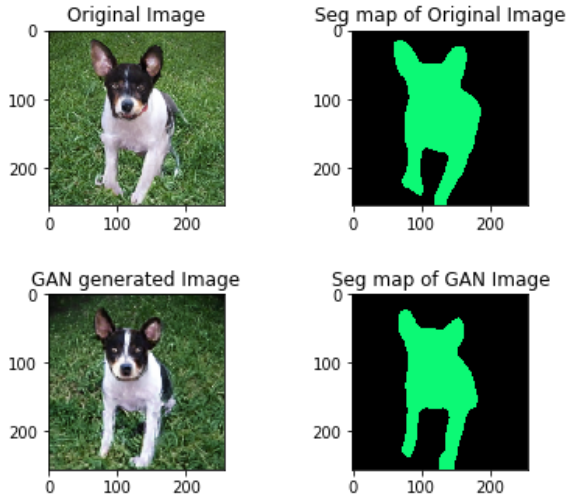

```
----------------------------------------------------------------------
----------------------------------
cost_old 4  is  tensor(1.3802, device='cuda:0', grad_fn=<SqrtBackw
ard>)
cost 4  is  tensor(0.1812, device='cuda:0', grad_fn=<NllLoss2DBack
ward>)
```



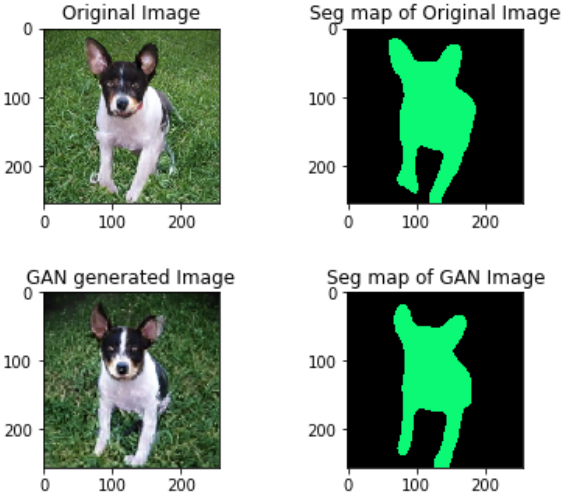

```
----------------------------------------------------------------------
----------------------------------
cost_old 5  is  tensor(1.4377, device='cuda:0', grad_fn=<SqrtBackw
ard>)
cost 5  is  tensor(0.1667, device='cuda:0', grad_fn=<NllLoss2DBack
ward>)
```



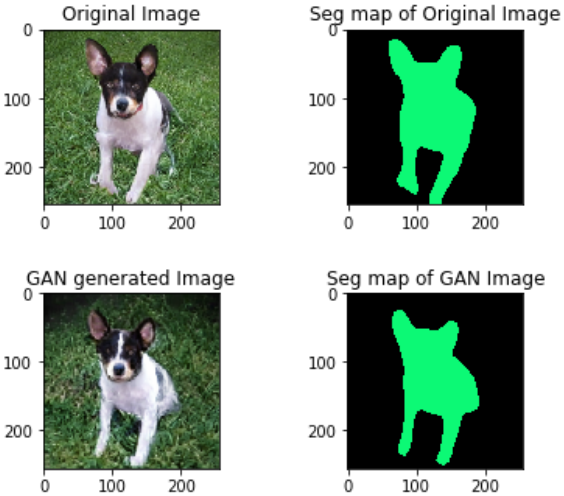

```
----------------------------------------------------------------------
----------------------------------
cost_old 6  is  tensor(1.2665, device='cuda:0', grad_fn=<SqrtBackw
ard>)
cost 6  is  tensor(0.1402, device='cuda:0', grad_fn=<NllLoss2DBack
ward>)
```



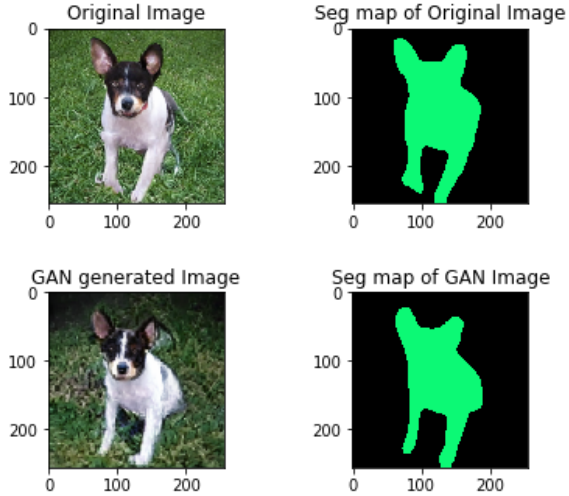
```
----------------------------------------------------------------------
----------------------------------
cost_old 7  is  tensor(1.3541, device='cuda:0', grad_fn=<SqrtBackw
ard>)
cost 7  is  tensor(0.1431, device='cuda:0', grad_fn=<NllLoss2DBack
ward>)
```



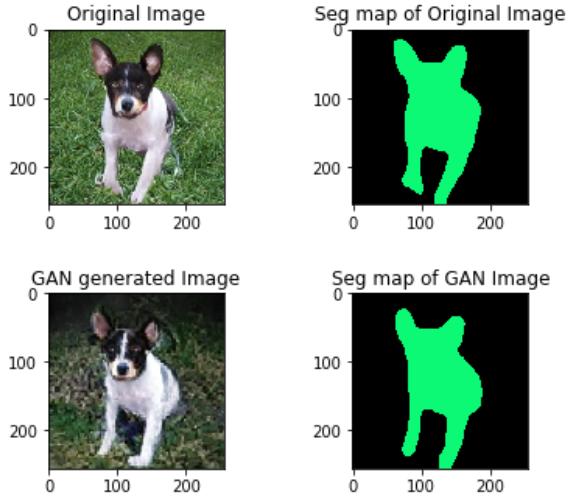

```
----------------------------------------------------------------------
----------------------------------
cost_old 8  is  tensor(1.2407, device='cuda:0', grad_fn=<SqrtBackw
ard>)
cost 8  is  tensor(0.1430, device='cuda:0', grad_fn=<NllLoss2DBack
ward>)
```



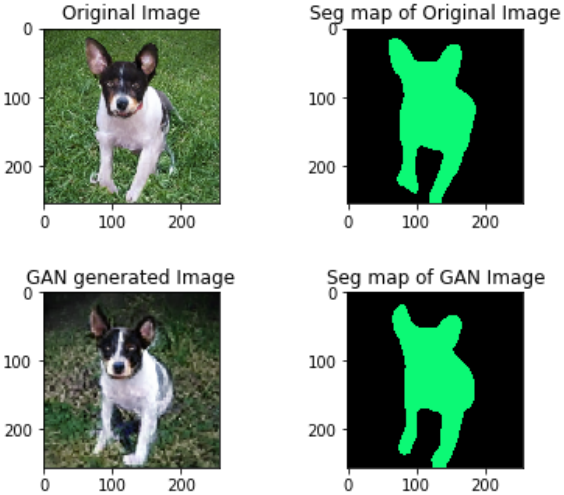

```
----------------------------------------------------------------------
----------------------------------
cost_old 9  is  tensor(1.7071, device='cuda:0', grad_fn=<SqrtBackw
ard>)
cost 9  is  tensor(0.1250, device='cuda:0', grad_fn=<NllLoss2DBack
ward>)
```



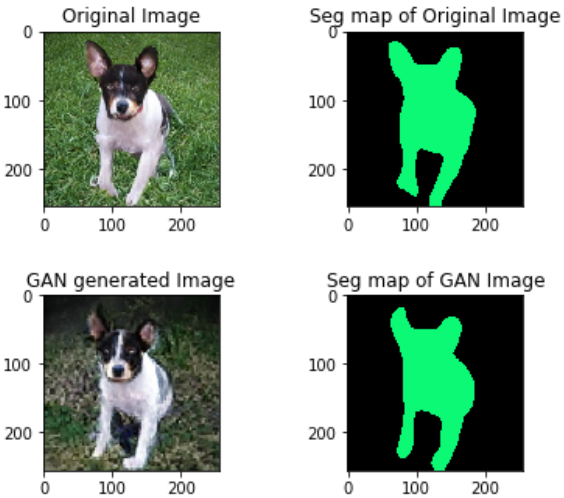

```
----------------------------------------------------------------------
----------------------------------
cost_old 10  is  tensor(1.6699, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 10  is  tensor(0.1099, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)
```



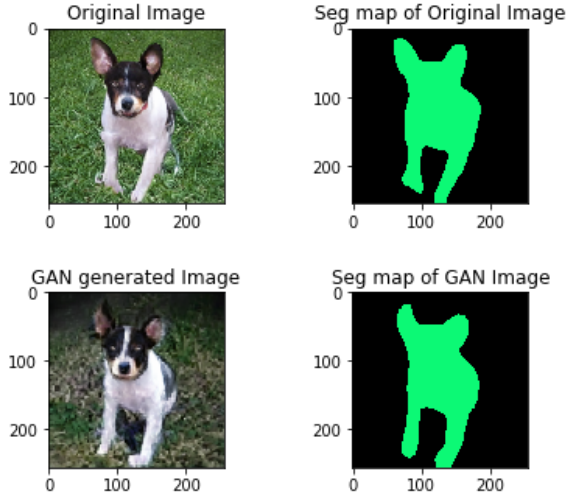
```
----------------------------------------------------------------------
----------------------------------
cost_old 11  is  tensor(1.6502, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 11  is  tensor(0.1084, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)
```



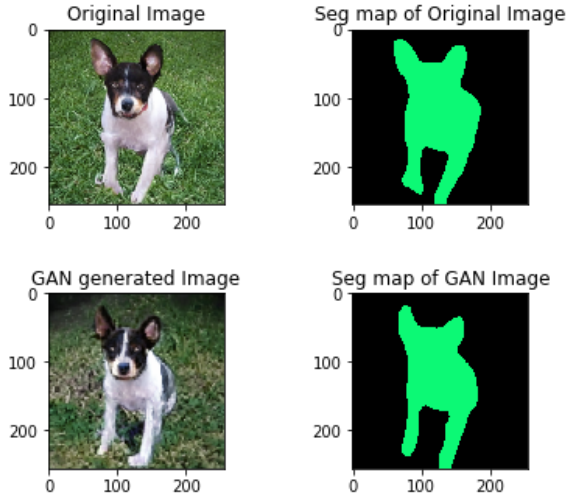

```
----------------------------------------------------------------------
----------------------------------
cost_old 12  is  tensor(1.6136, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 12  is  tensor(0.0967, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)
```



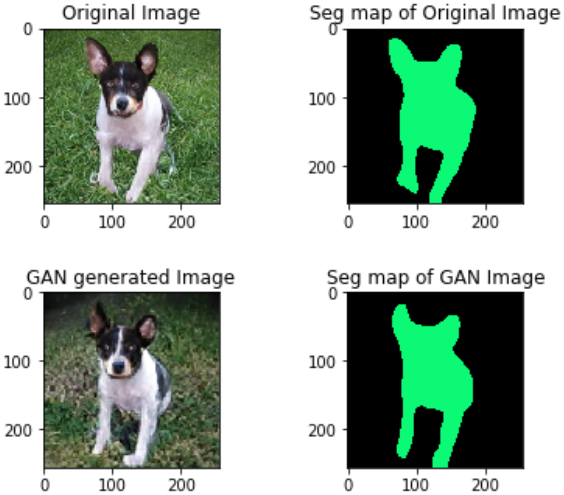

```
----------------------------------------------------------------------
----------------------------------
cost_old 13  is  tensor(1.6040, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 13  is  tensor(0.1011, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)
```



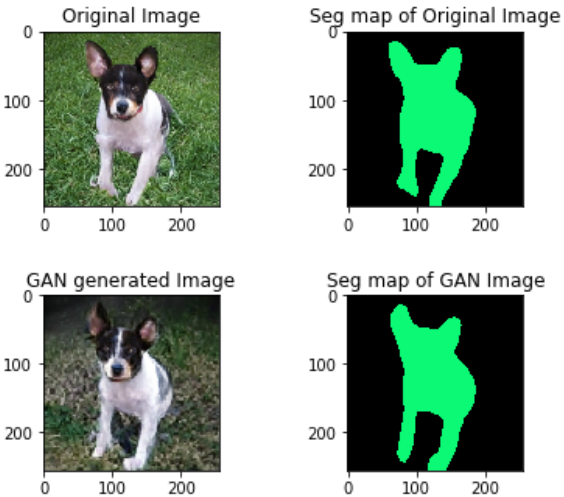

```
----------------------------------------------------------------------
----------------------------------
cost_old 14  is  tensor(1.6727, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 14  is  tensor(0.1088, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)
```



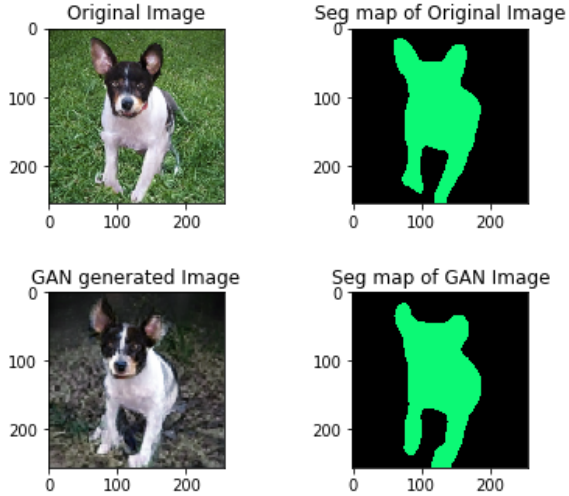
```
----------------------------------------------------------------------
----------------------------------
cost_old 15  is  tensor(1.5295, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 15  is  tensor(0.1020, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)
```



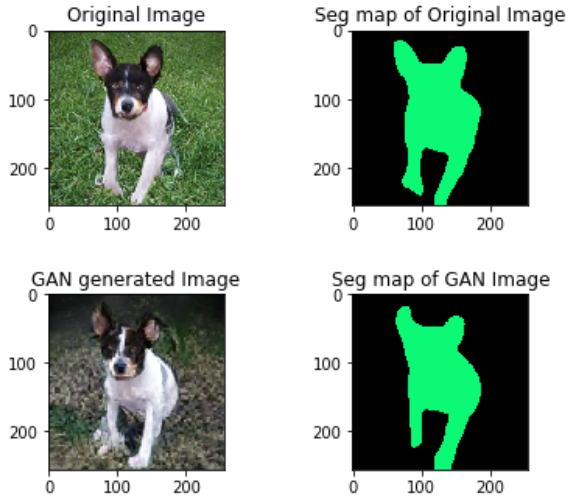

```
----------------------------------------------------------------------
----------------------------------
cost_old 16  is  tensor(1.5759, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 16  is  tensor(0.0936, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)
```



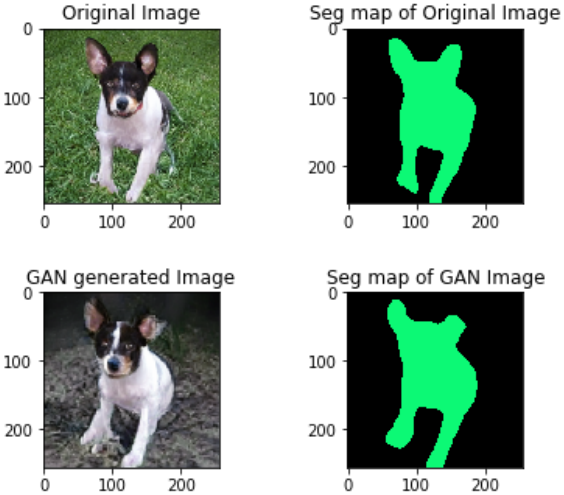

```
----------------------------------------------------------------------
----------------------------------
cost_old 17  is  tensor(1.6073, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 17  is  tensor(0.0953, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)
```




```
----------------------------------------------------------------------
----------------------------------
cost_old 18  is  tensor(1.6425, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 18  is  tensor(0.0966, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)
```



```
----------------------------------------------------------------------
----------------------------------
cost_old 19  is  tensor(1.3843, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 19  is  tensor(0.1226, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)
```




```
----------------------------------------------------------------------
----------------------------------
cost_old 20  is  tensor(1.3282, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 20  is  tensor(0.1183, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)
```