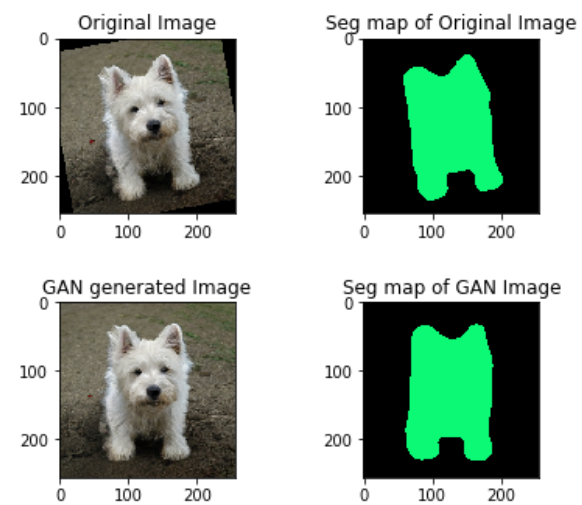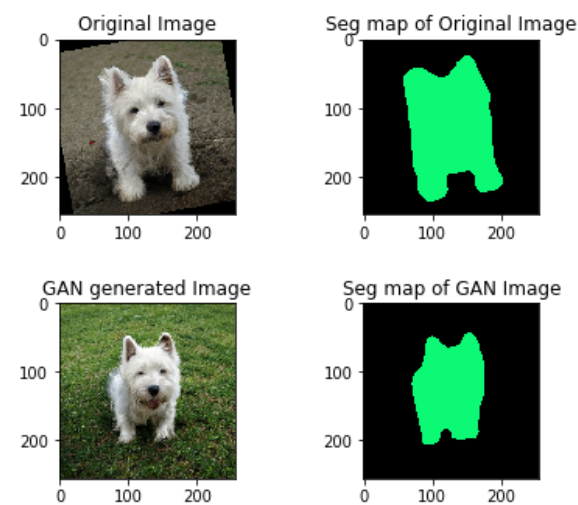```
    print('-'*100)
    cost.backward()
    optimizer.step()

print('End')
```
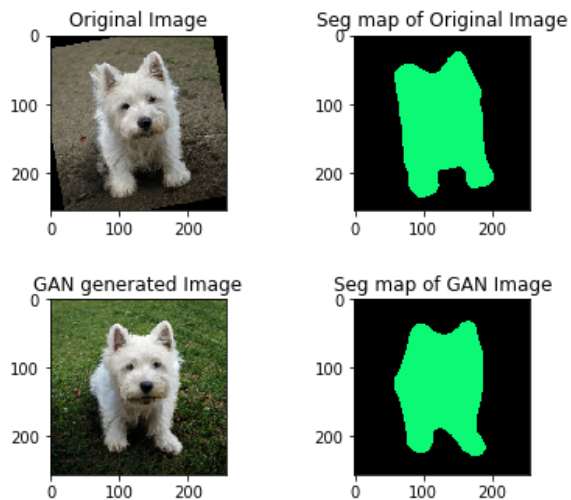
```
cost_old 1 is tensor(1.3727, device='cuda:0', grad_fn=<SqrtBackw
ard>)
cost 1 is tensor(0.2020, device='cuda:0', grad_fn=<NllLoss2DBack
ward>)
```



```

----
cost_old 2 is tensor(1.5648, device='cuda:0', grad_fn=<SqrtBackw
ard>)
cost 2 is tensor(0.5903, device='cuda:0', grad_fn=<NllLoss2DBack
ward>)
```
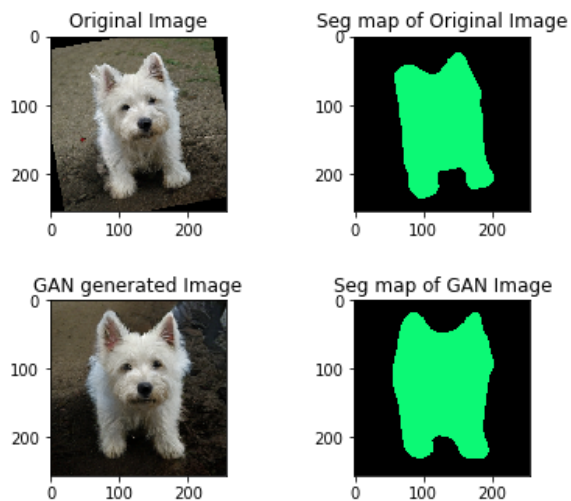
------------------------------------------------------------------
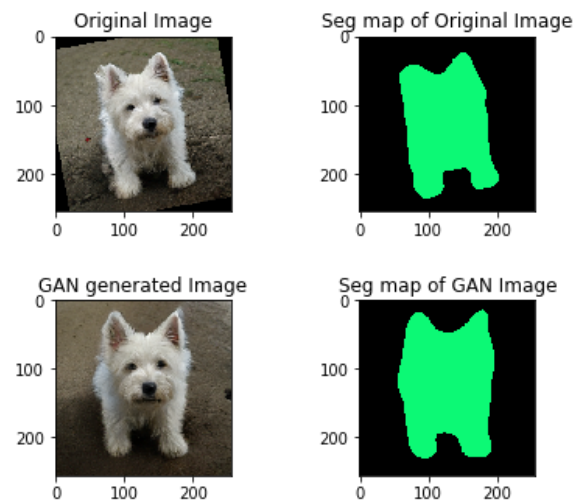-----------------------------------

cost_old 3  is  tensor(1.4038, device='cuda:0', grad_fn=<SqrtBackward>)

cost 3  is  tensor(0.2287, device='cuda:0', grad_fn=<NllLoss2DBackward>)



------------------------------------------------------------------
-----------------------------------
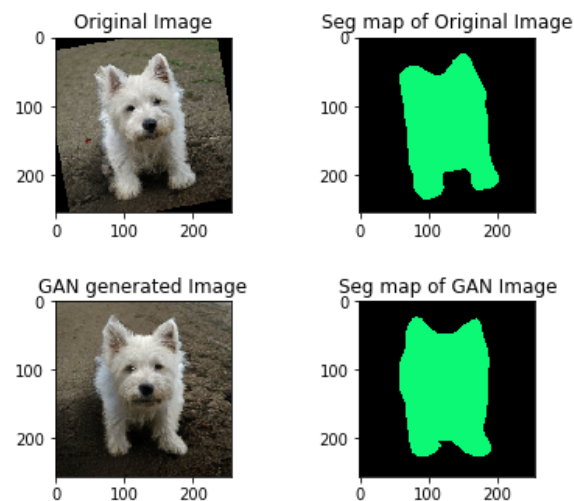
cost_old 4  is  tensor(1.2215, device='cuda:0', grad_fn=<SqrtBackward>)

cost 4  is  tensor(0.2670, device='cuda:0', grad_fn=<NllLoss2DBackward>)



------------------------------------------------------------------
-----------------------------------

cost_old 5  is  tensor(1.2045, device='cuda:0', grad_fn=<SqrtBackward>)

cost 5  is  tensor(0.2462, device='cuda:0', grad_fn=<NllLoss2DBackward>)



------------------------------------------------------------------
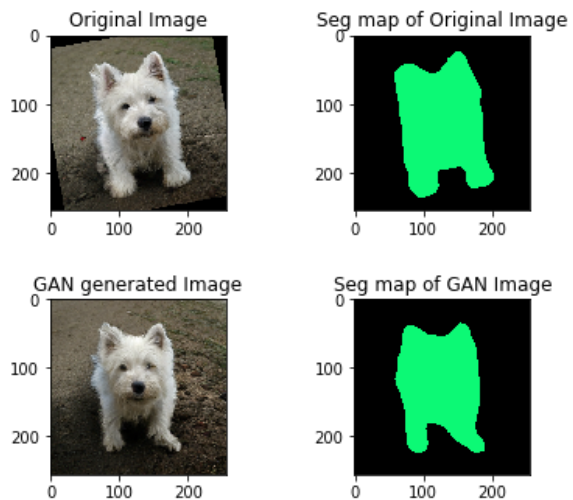-----------------------------------

cost_old 6  is  tensor(1.2642, device='cuda:0', grad_fn=<SqrtBackward>)

cost 6  is  tensor(0.1794, device='cuda:0', grad_fn=<NllLoss2DBackward>)

------------------------------------------------------------------
-----------------------------------

cost_old 7  is  tensor(1.2037, device='cuda:0', grad_fn=<SqrtBackward>)
cost 7  is  tensor(0.1651, device='cuda:0', grad_fn=<NllLoss2DBackward>)



------------------------------------------------------------------
-----------------------------------

cost_old 8  is  tensor(1.1951, device='cuda:0', grad_fn=<SqrtBackward>)
cost 8  is  tensor(0.1318, device='cuda:0', grad_fn=<NllLoss2DBackward>)



------------------------------------------------------------------
-----------------------------------
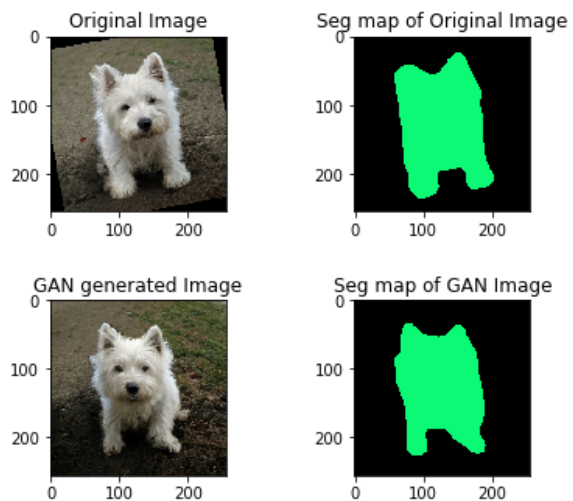
cost_old 9  is  tensor(1.2265, device='cuda:0', grad_fn=<SqrtBackward>)
cost 9  is  tensor(0.1320, device='cuda:0', grad_fn=<NllLoss2DBackward>)



------------------------------------------------------------------
-----------------------------------
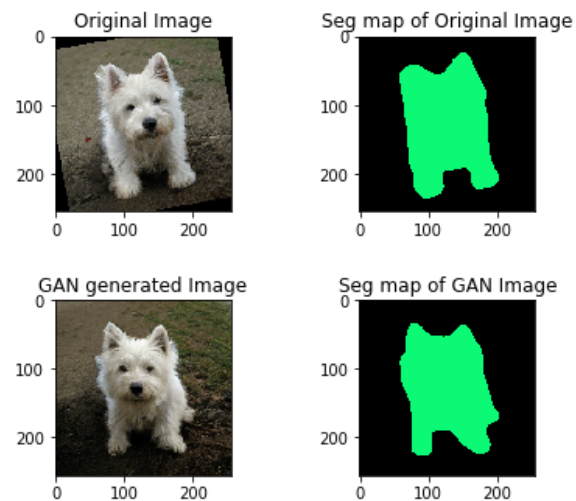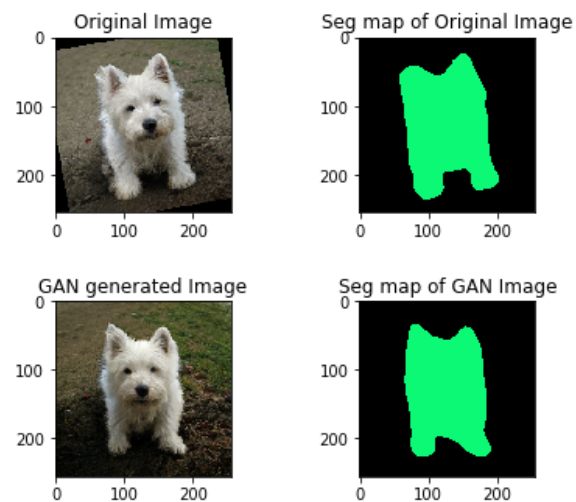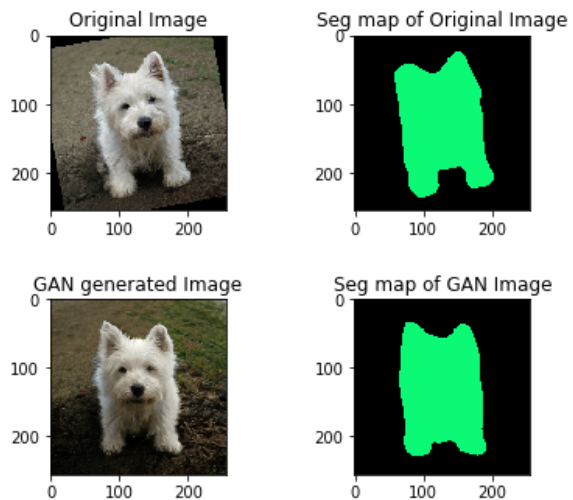
cost_old 10  is  tensor(1.3930, device='cuda:0', grad_fn=<SqrtBackward>)
cost 10  is  tensor(0.1451, device='cuda:0', grad_fn=<NllLoss2DBackward>)

------------------------------------------------------------------
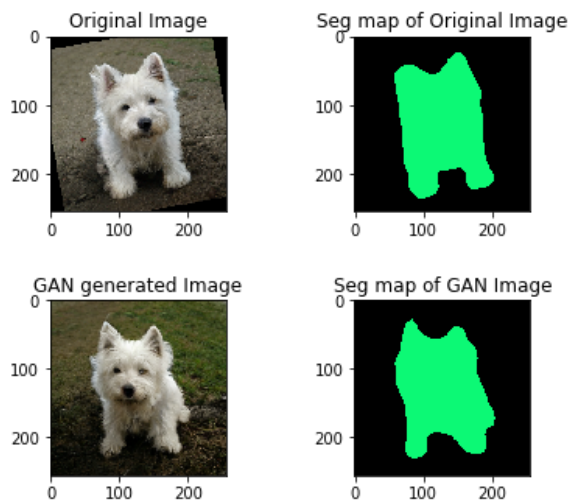-----------------------------------
cost_old 11  is  tensor(1.4127, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 11  is  tensor(0.1401, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)



------------------------------------------------------------------
-----------------------------------
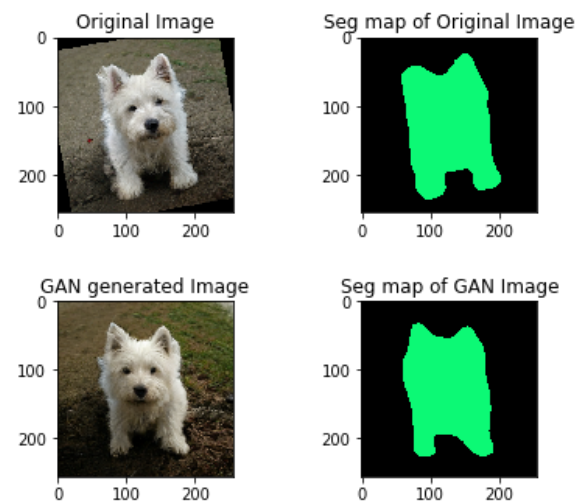cost_old 12  is  tensor(1.2539, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 12  is  tensor(0.1406, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)

------------------------------------------------------------------
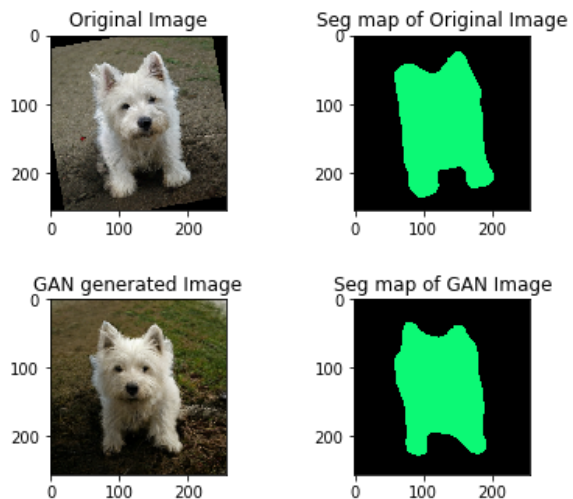-----------------------------------
cost_old 13  is  tensor(1.1711, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 13  is  tensor(0.1314, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)
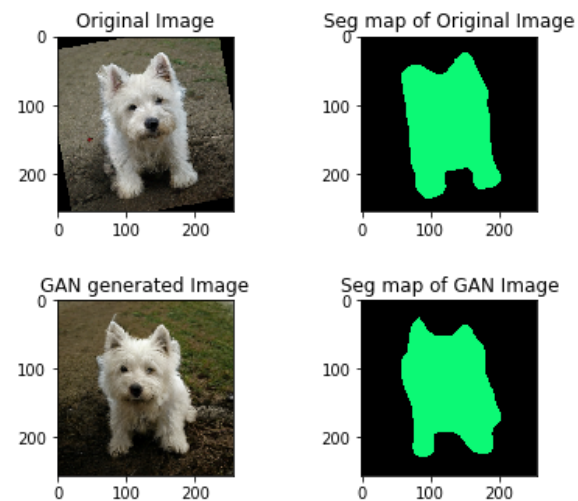


------------------------------------------------------------------
-----------------------------------
cost_old 14  is  tensor(1.3578, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 14  is  tensor(0.1202, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)

------------------------------------------------------------------
-----------------------------------
cost_old 15  is  tensor(1.2777, device='cuda:0', grad_fn=<SqrtBackward>)
cost 15  is  tensor(0.1155, device='cuda:0', grad_fn=<NllLoss2DBackward>)



------------------------------------------------------------------
-----------------------------------
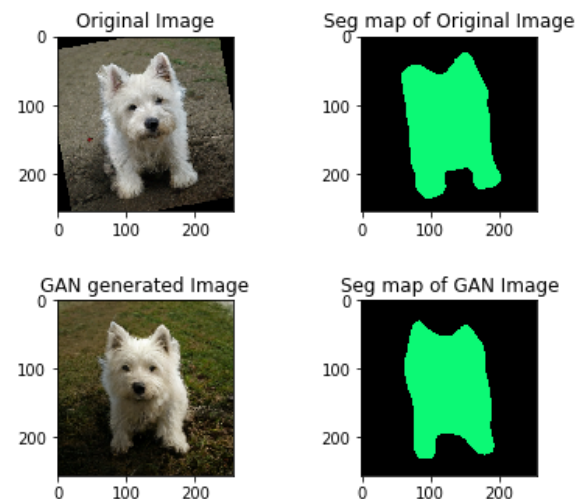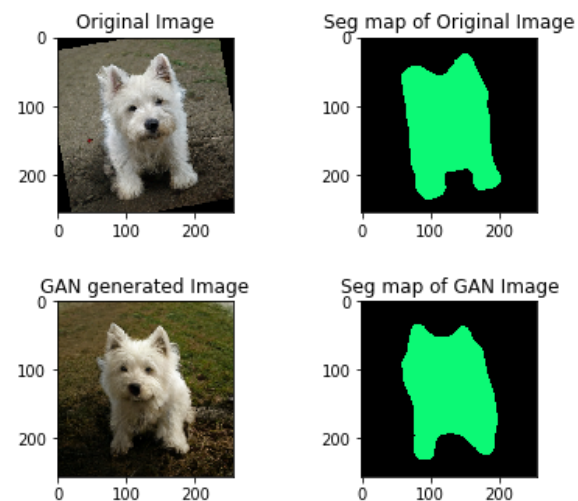cost_old 16  is  tensor(1.3238, device='cuda:0', grad_fn=<SqrtBackward>)
cost 16  is  tensor(0.1167, device='cuda:0', grad_fn=<NllLoss2DBackward>)
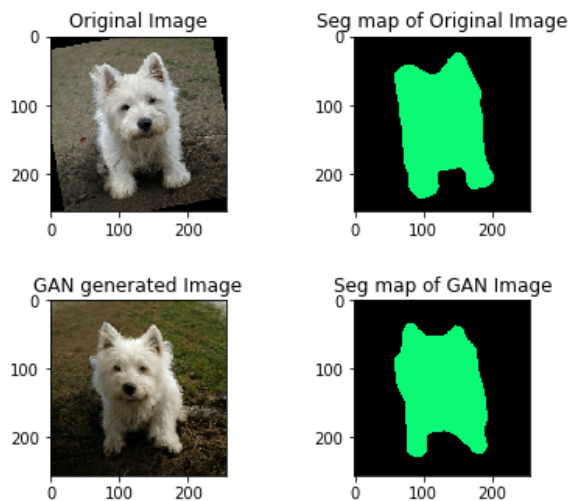
------------------------------------------------------------------
-----------------------------------
cost_old 17  is  tensor(1.4213, device='cuda:0', grad_fn=<SqrtBackward>)
cost 17  is  tensor(0.1080, device='cuda:0', grad_fn=<NllLoss2DBackward>)



------------------------------------------------------------------
-----------------------------------
cost_old 18  is  tensor(1.3601, device='cuda:0', grad_fn=<SqrtBackward>)
cost 18  is  tensor(0.1143, device='cuda:0', grad_fn=<NllLoss2DBackward>)

```

----
cost_old 19 is tensor(1.3316, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 19 is tensor(0.1103, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)
```
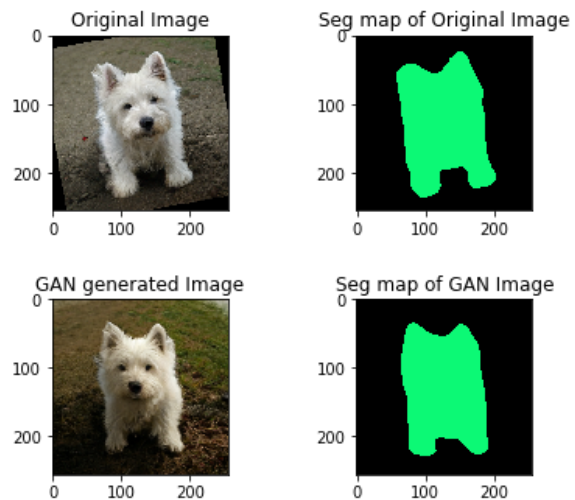


```
----------------------------------------------------------------
-----------------------------------
cost_old 20  is  tensor(1.2354, device='cuda:0', grad_fn=<SqrtBack
ward>)
cost 20  is  tensor(0.1092, device='cuda:0', grad_fn=<NllLoss2DBac
kward>)
```