

Encoder Architecture Detailed Summary

Encoder (Downsampling) Blocks

Layer	Input Dims	Output Dims	Filters	Filter Shape	Stride	Padding	Normalization
down1	256x256x3	128x128x64	64	4x4	2	1	No
down2	128x128x64	64x64x128	128	4x4	2	1	Yes
down3	64x64x128	32x32x256	256	4x4	2	1	Yes
down4	32x32x256	16x16x512	512	4x4	2	1	Yes
down5	16x16x512	8x8x512	512	4x4	2	1	Yes
down6	8x8x512	4x4x512	512	4x4	2	1	Yes
down7	4x4x512	2x2x512	512	4x4	2	1	No

Decoder (Upsampling) Blocks

Layer	Input Dims	Output Dims	Concat	Total Channels	Filters	Filter Shape	Stride	Padding
up1	2x2x512	4x4x512	None	512	512	4x4	2	1
up2	4x4x512	8x8x512	4x4x512 (down6)	1024	512	4x4	2	1
up3	8x8x512	16x16x512	8x8x512(down5)	1024	512	4x4	2	1
up4	16x16x512	32x32x256	16x16x512(down4)	1024	256	4x4	2	1
up5	32x32x256	64x64x128	32x32x256(down3)	512	128	4x4	2	1
up6	64x64x128	128x128x64	64x64x128(down2)	256	64	4x4	2	1
final	128x128x64	256x256x3	128x128x64(down2)	128	3	4x4	2	1