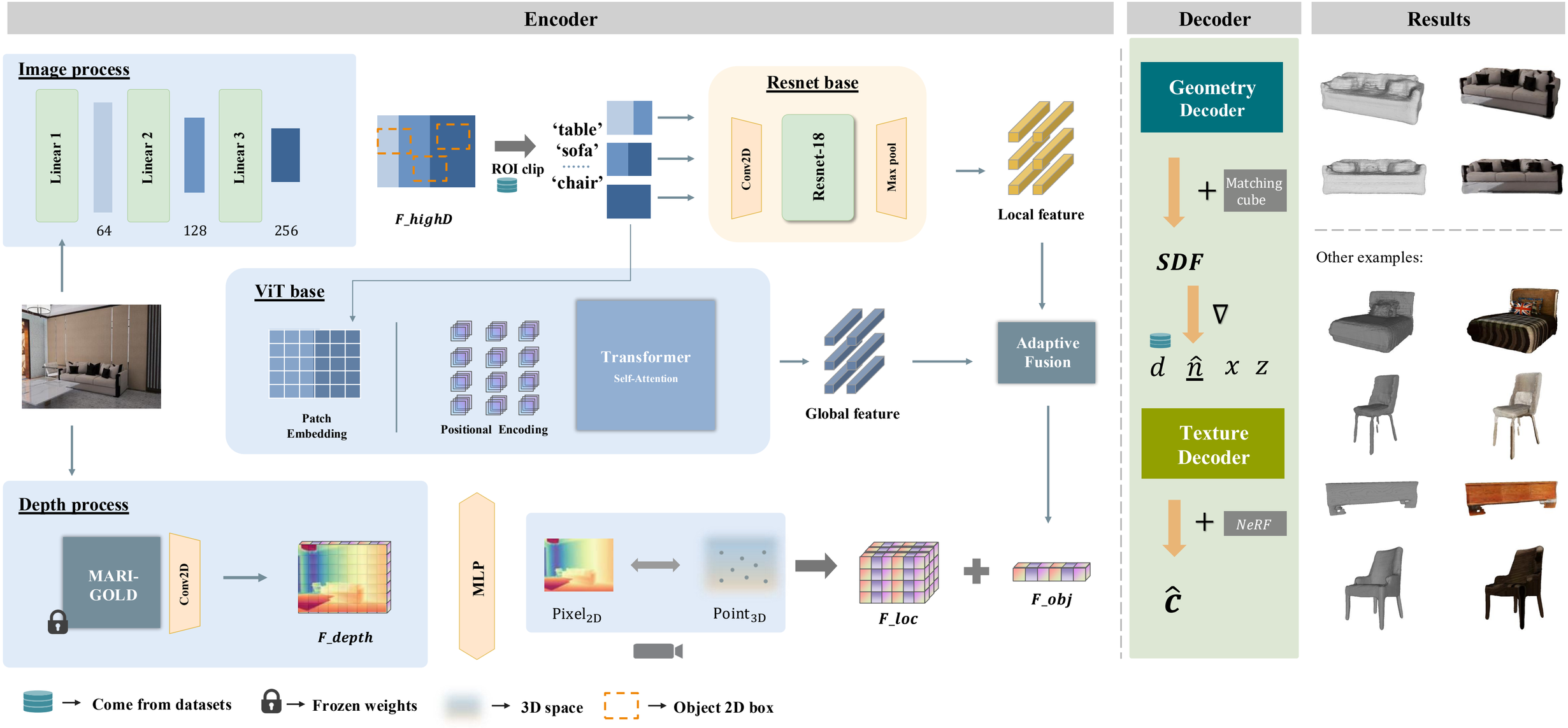
Encoder
Decoder
Results
Image process
Linear 1
Linear 2
Linear 3
64
128
256
F_highD
ROI clip
'table'
'sofa'
'chair'
Resnet base
Conv2D
Resnet-18
Max pool
Local feature
ViT base
Patch Embedding
Positional Encoding
Transformer
Self-Attention
Global feature
Adaptive Fusion
Depth process
MARI-GOLD
Conv2D
F_depth
MLP
Pixel2D
Point3D
F_loc
F_obj
Geometry Decoder
Matching cube
SDF
∇
d n̂ x z
Texture Decoder
NeRF
ĉ
Other examples:
Come from datasets
Frozen weights
3D space
Object 2D box