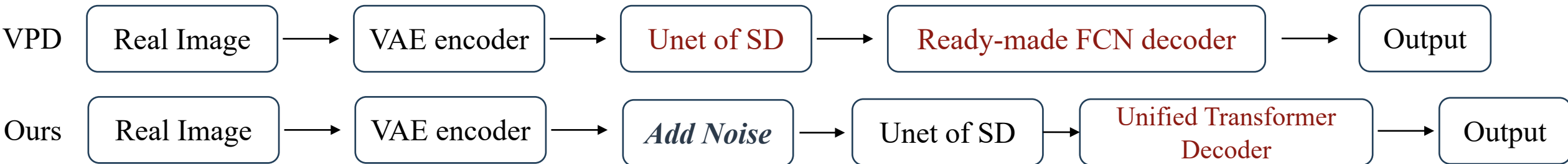
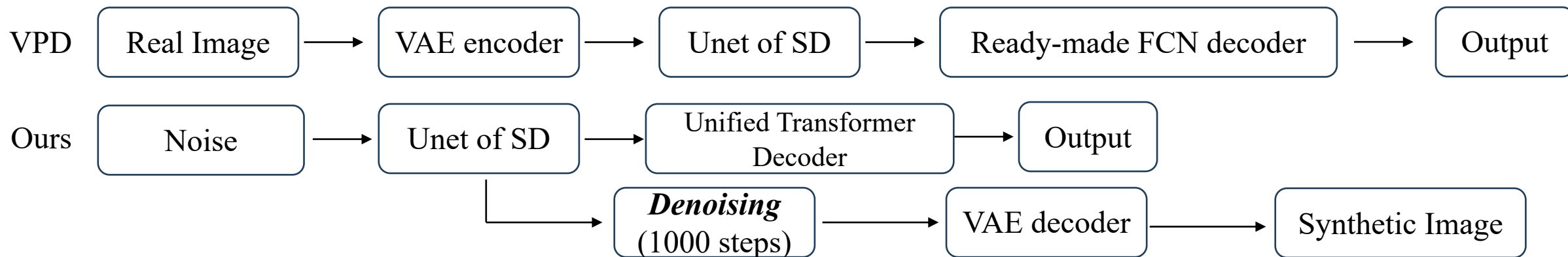


## Training

In **color** denotes trainable weight.



## Inference



Conclusion: VPD is a perception model (similar to Faster R-CNN), while our DatasetDM is a generative model. The most significant difference is that VPD lacks the **diffusion process**, which results in distinct architectures, principles, and training approaches for the two.