



Input Image



**Large Vision-Language Model
as Feature Extractor**



Discriminator