

Figure R1: Visual results for different content extractions on ControlNet or our proposed content-fusion encoder (\*).

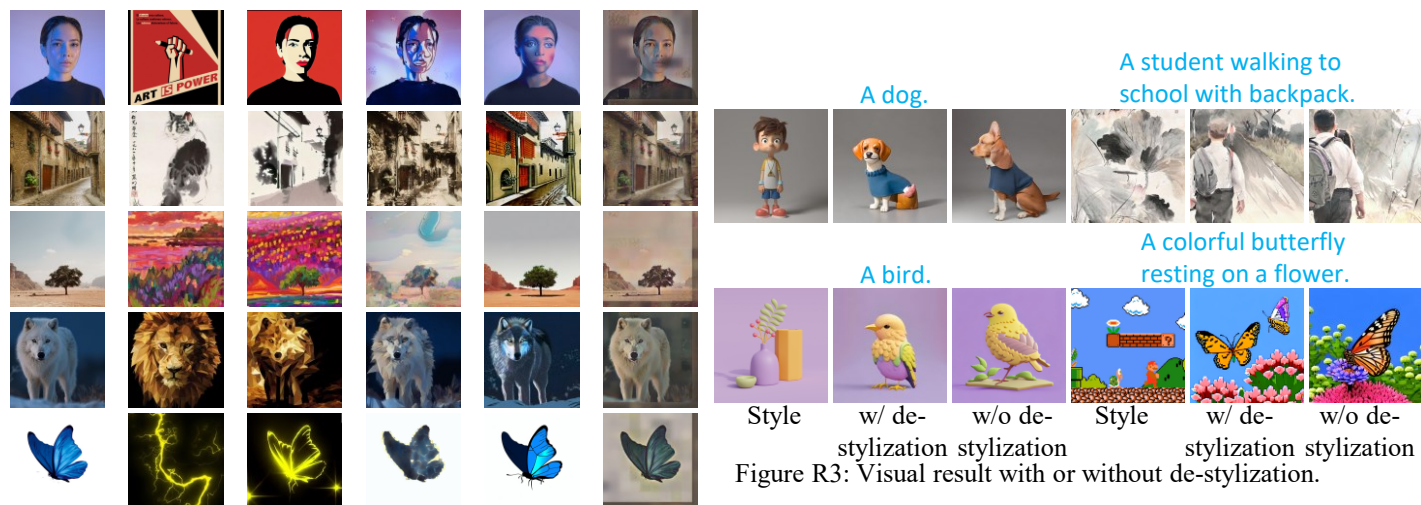


Figure R3: Visual result with or without de-stylization.

Figure R2: Visual comparison among StyleShot, DreamStyler, Z-Star and Master.



Figure R4: Qualitative comparison with SOTA text-driven style transfer methods on third-party bench from [1].

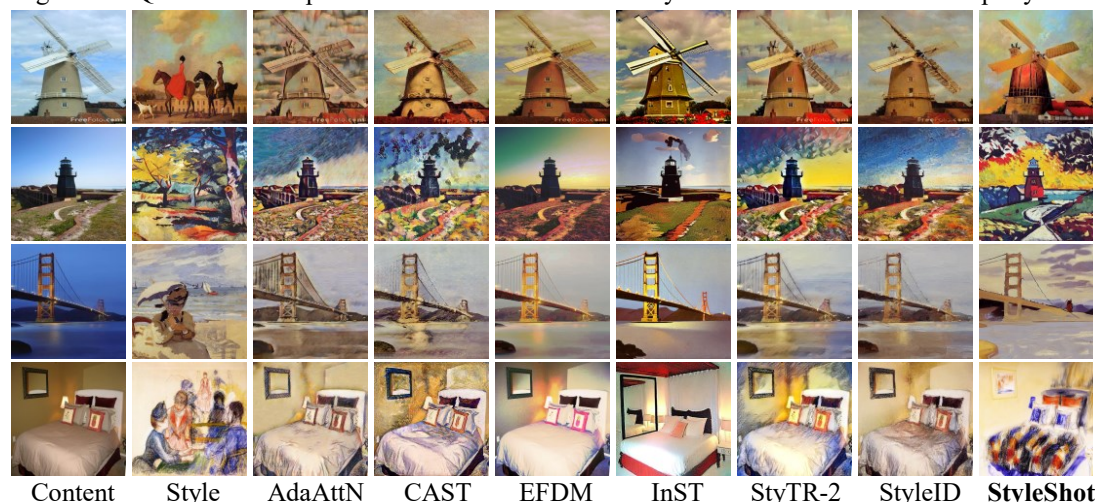


Figure R5: Qualitative comparison with SOTA image-driven style transfer methods on third-party bench from [1].

[1]. "A Comprehensive Evaluation of Arbitrary Image Style Transfer Methods", IEEE TVCG.