I. Inducing Context Tokens with Style Tokens

G

AS-Net

Residual Block

Stem

CNN Encoder

Style Patches

Meta-Net

v_1 v_2 ... v_n

Style Tokens

Context Tokens

Linear Projection

Text (C)

"Said"

D → L_D

Mixed Tokens

II. Capture Dependencies

Style Samples (X_w)

MOVE
Foot
more
Labour

Reshape & Flatten

F_s

Transformer Encoder

F_{st}

Transformer Decoder

Up Sampling Block

Fake Samples (Y_w^C)

Said

W → L_W

R → L_R

G Style-induced Context-aware Generator $\quad$ D Discriminator $\quad$ R Recognizer $\quad$ W Writer Classifier $\quad$ F_s Style feature sequence $\quad$ F_{st} Style-text entanglement