AImage Lab

UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Handwritten Text Generation from Visual Archetypes

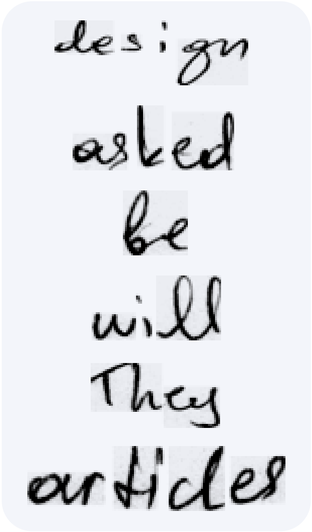
Vittorio Pippi, Silvia Cascianelli, Rita Cucchiara

June 22, 2023 THU-PM-177

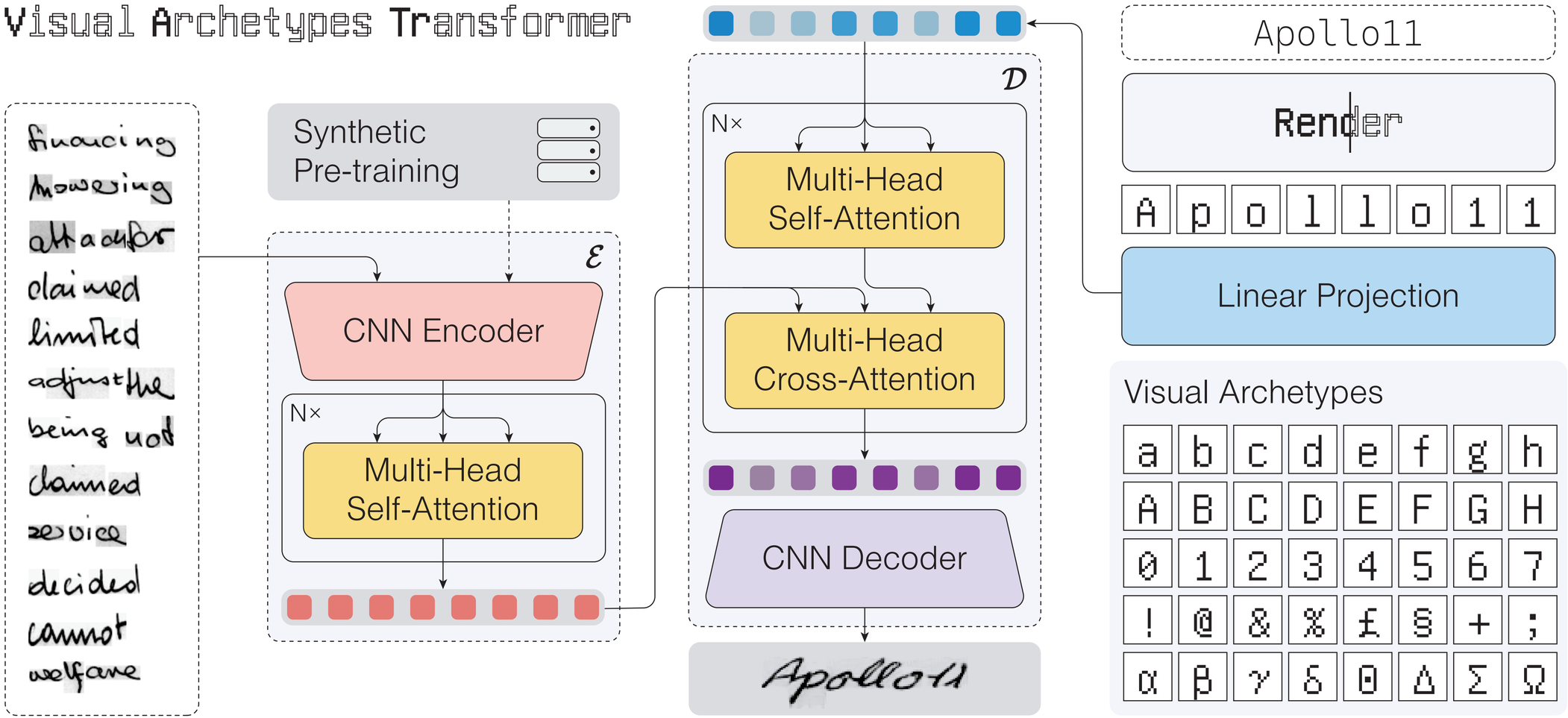
Unseen words

planets	lips	varies
impact	skips	Gold
maple	voyager	noisy
stick	forums	drafts

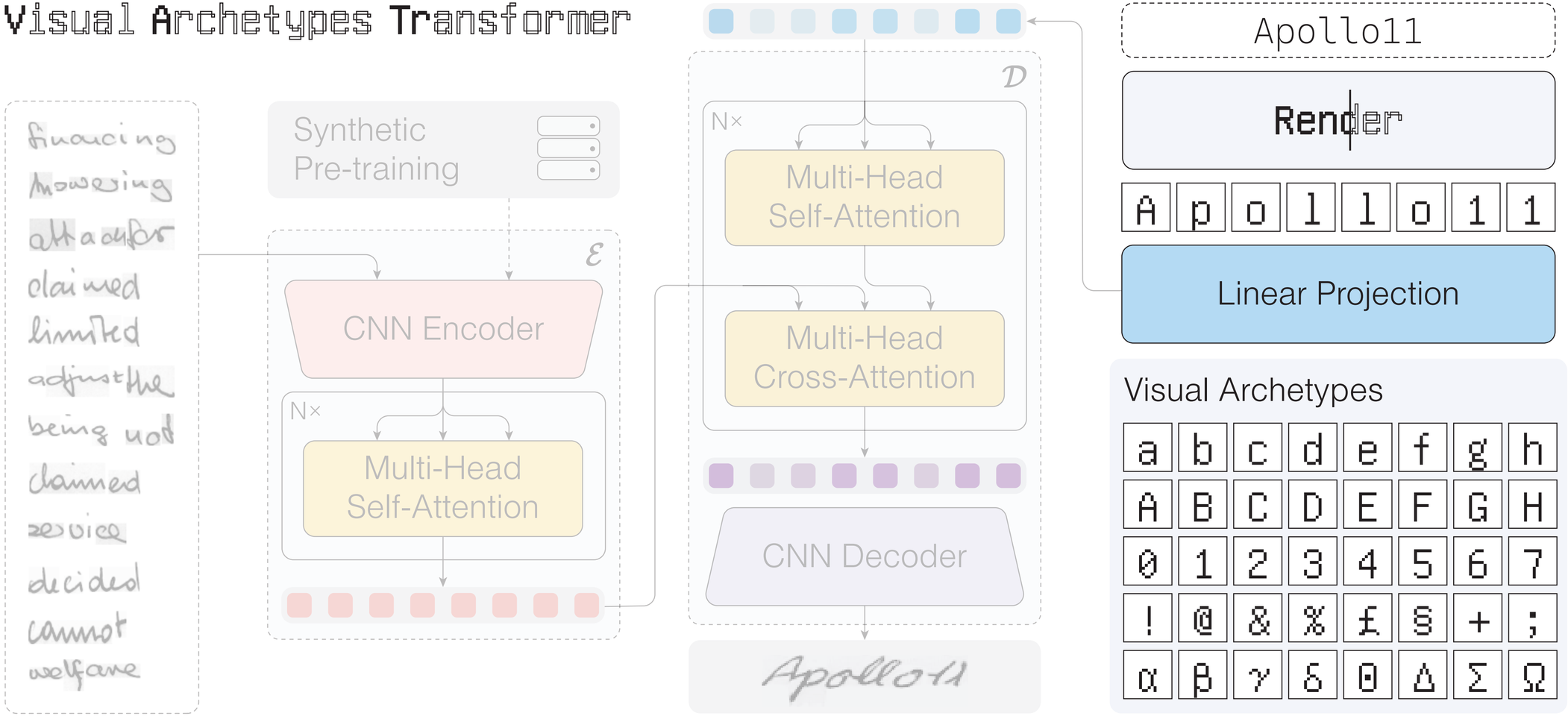
Unseen styles



Axis	lips	jazz	query	voyager	Noisy
Next	skips	Gold	stick	9th	drafts

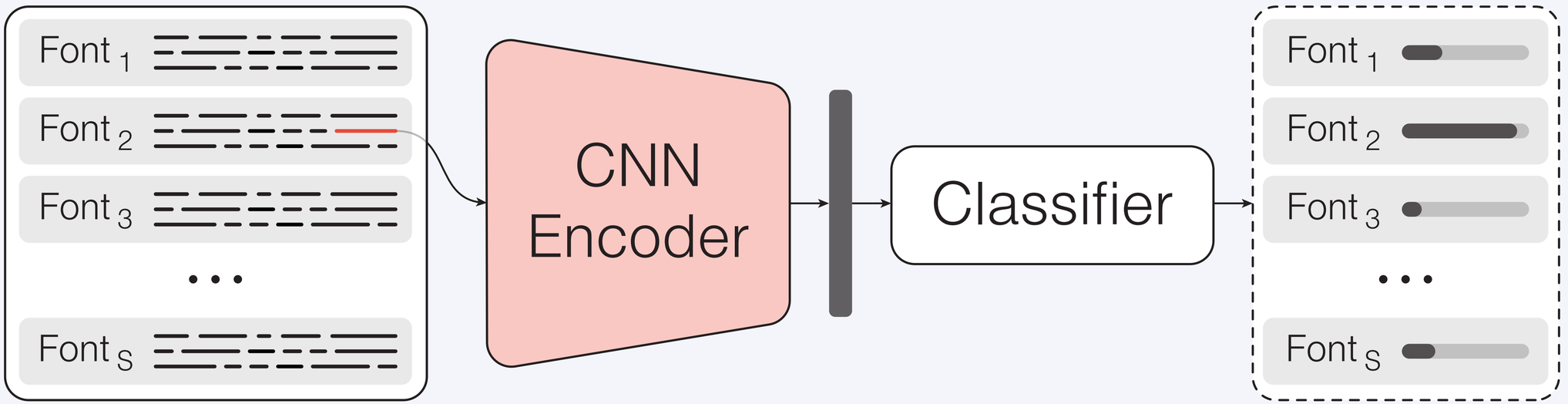

Visual Archetypes Transformer
Synthetic
Pre-training
$\mathcal{E}$
CNN Encoder
N×
Multi-Head
Self-Attention
$\mathcal{D}$
N×
Multi-Head
Self-Attention
Multi-Head
Cross-Attention
CNN Decoder
Apollo11
Apollo11
Render
A p o l l o 1 1
Linear Projection
Visual Archetypes
a b c d e f g h
A B C D E F G H
0 1 2 3 4 5 6 7
! @ & % £ § + ;
α β γ δ Θ Δ Σ Ω

Visual Archetypes Transformer



Synthetic Pre-training

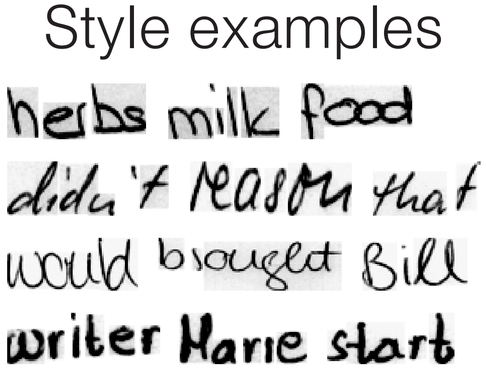
The convolutional encoder is trained to perform style classification on a synthetic dataset capturing high variability.



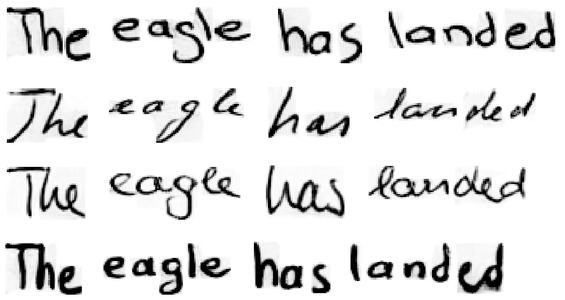
	ALL	IV-S	IV-U	OOV-S	OOV-U
TS-GAN	20.65	118.56	128.75	127.11	136.67
GANwriting	20.55	120.07	124.30	125.87	130.68
HWT	19.40	106.97	108.84	109.45	114.10
VATr (Ours)	**17.79**	**88.20**	**91.11**	**98.57**	**102.22**

	Long-Tail		Digits	
	FID	CER	FID	CER
HiGAN	26.08	**8.63**	26.08	101.53
HWT	40.95	20.36	40.95	98.47
VATr (Ours)	**21.36**	11.85	**21.36**	**94.66**

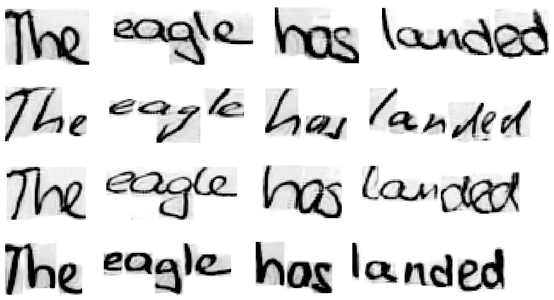
Style examples

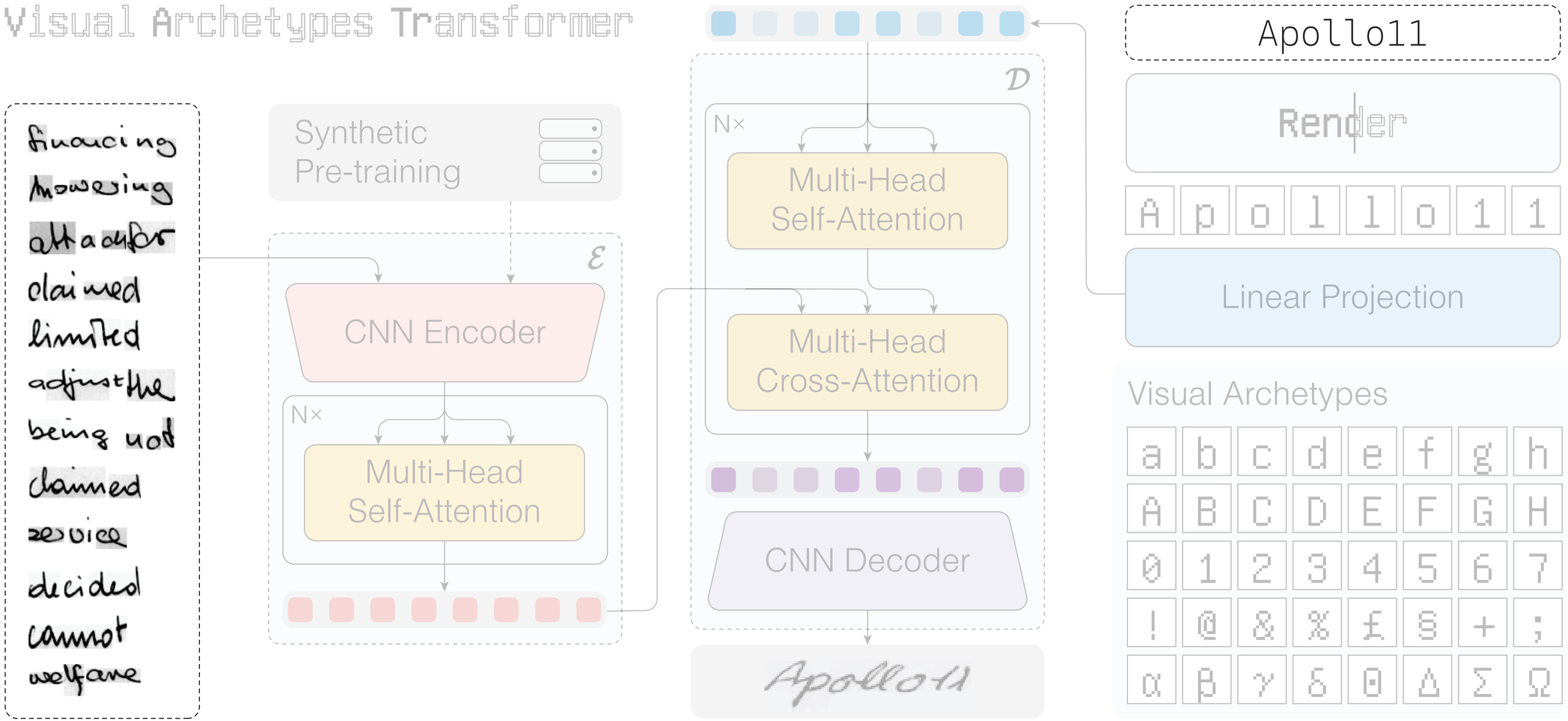


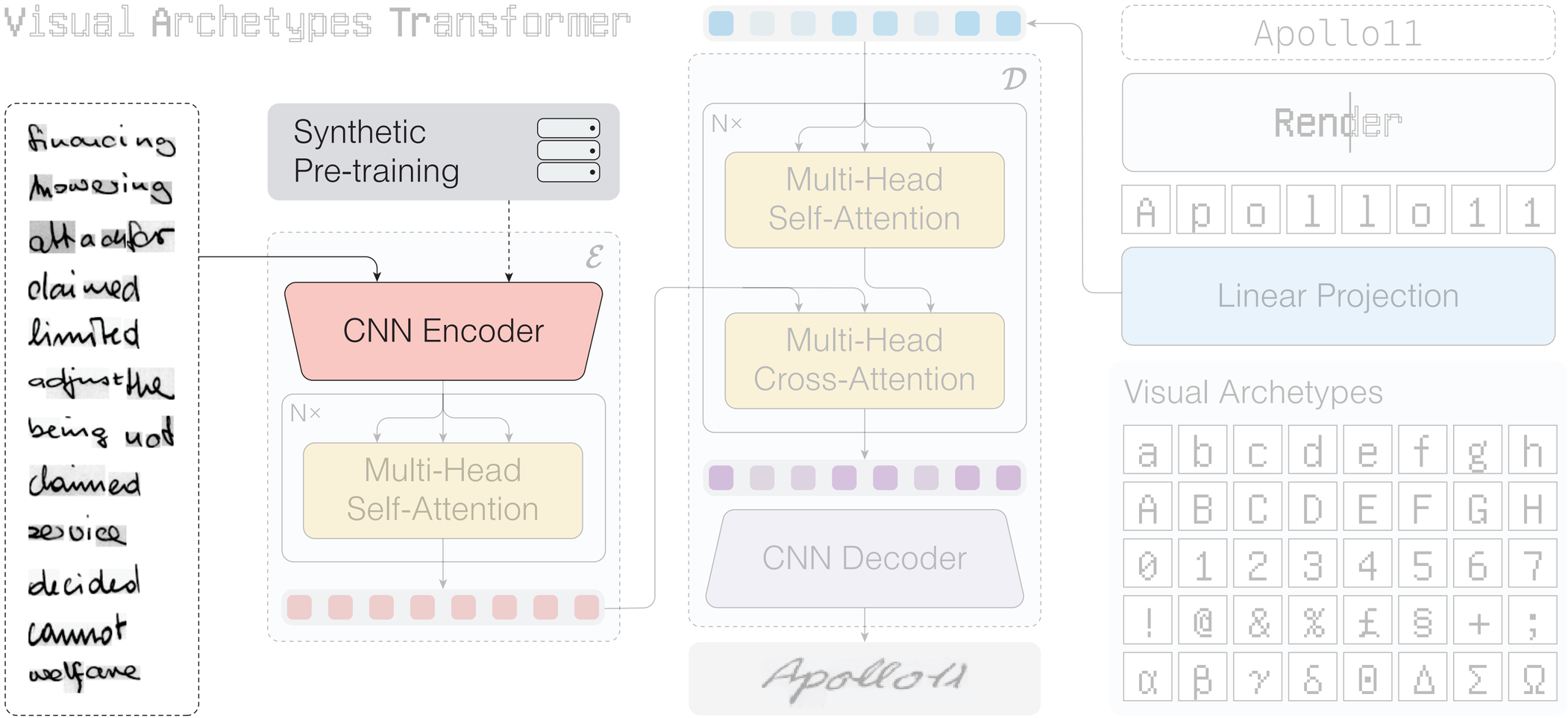
VATr



HWT




Visual Archetypes Transformer
financing
mowering
attacker
claimed
limited
adjust the
being not
claimed
service
decided
cannot
welfare
Synthetic
Pre-training
ℰ
CNN Encoder
N×
Multi-Head
Self-Attention
𝒟
N×
Multi-Head
Self-Attention
Multi-Head
Cross-Attention
CNN Decoder
Apollo11
Apollo11
Render
A p o l l o 1 1
Linear Projection
Visual Archetypes
a b c d e f g h
A B C D E F G H
0 1 2 3 4 5 6 7
! @ & % £ § + ;
α β γ δ θ Δ Σ Ω

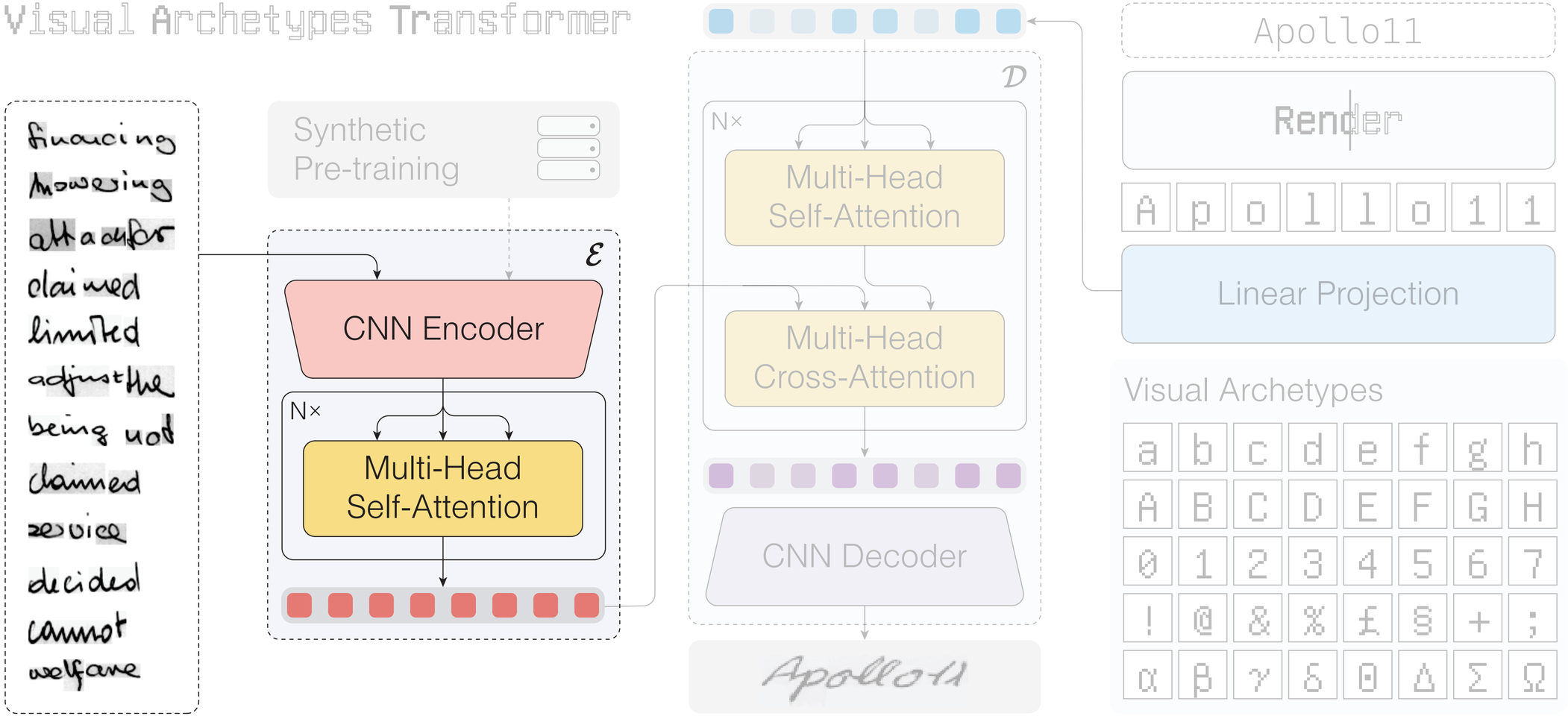

Visual Archetypes Transformer
financing
Mowering
attacker
claimed
limited
adjust the
being not
claimed
service
decided
cannot
welfare
Synthetic
Pre-training
ℰ
CNN Encoder
N×
Multi-Head
Self-Attention
𝒟
N×
Multi-Head
Self-Attention
Multi-Head
Cross-Attention
CNN Decoder
Apollo11
Apollo11
Render
A p o l l o 1 1
Linear Projection
Visual Archetypes
a b c d e f g h
A B C D E F G H
0 1 2 3 4 5 6 7
! @ & % £ § + ;
α β γ δ θ Δ Σ Ω

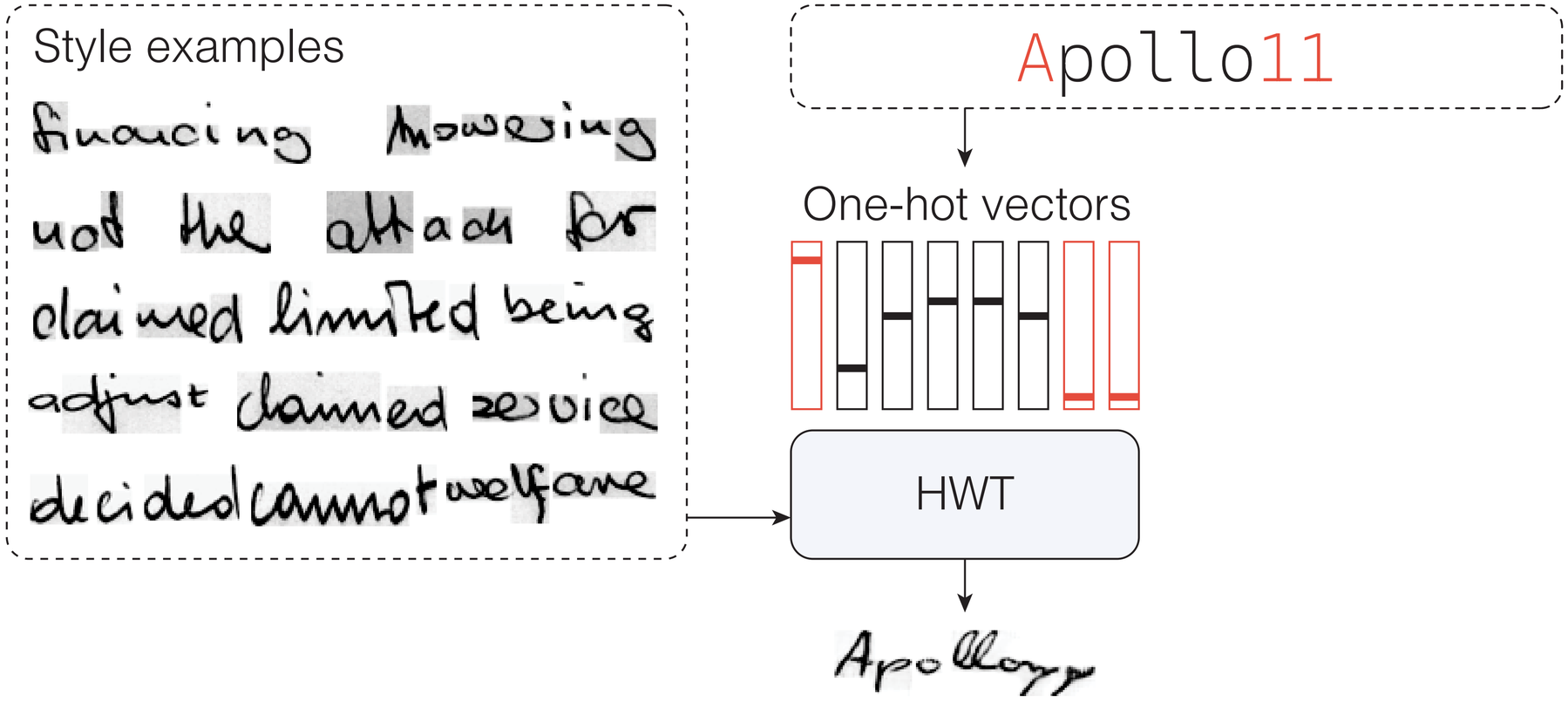
Words

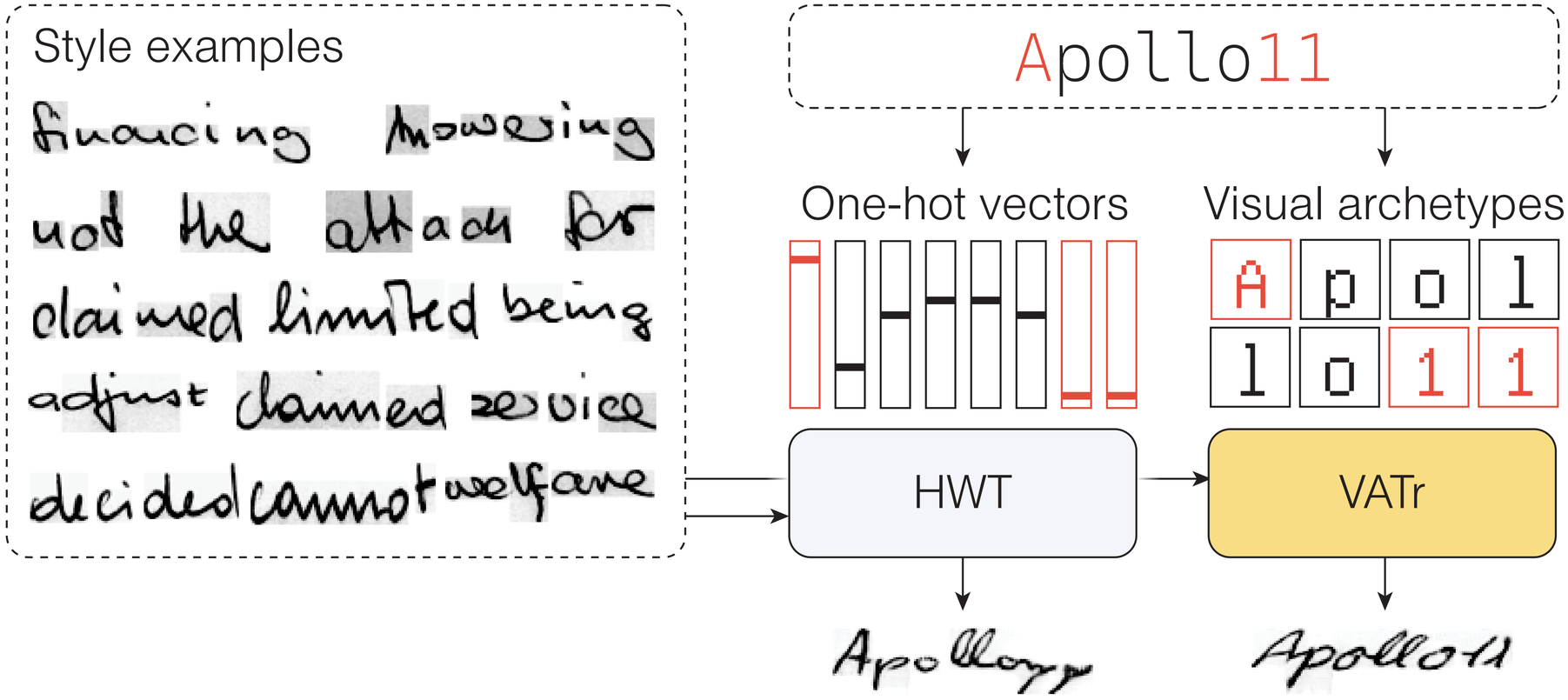
Fonts

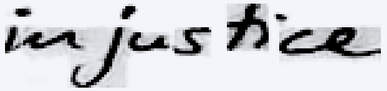
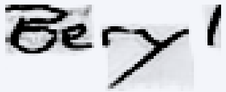
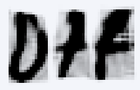
Apollo nebulae interstellar cosmic ... Hesperus
Apollo nebulae interstellar cosmic ... Hesperus
Apollo nebulae interstellar cosmic ... Hesperus
Apollo nebulae interstellar cosmic ... Hesperus
Apollo nebulae interstellar cosmic ... Hesperus
... ...
Apollo nebulae interstellar cosmic ... Hesperus

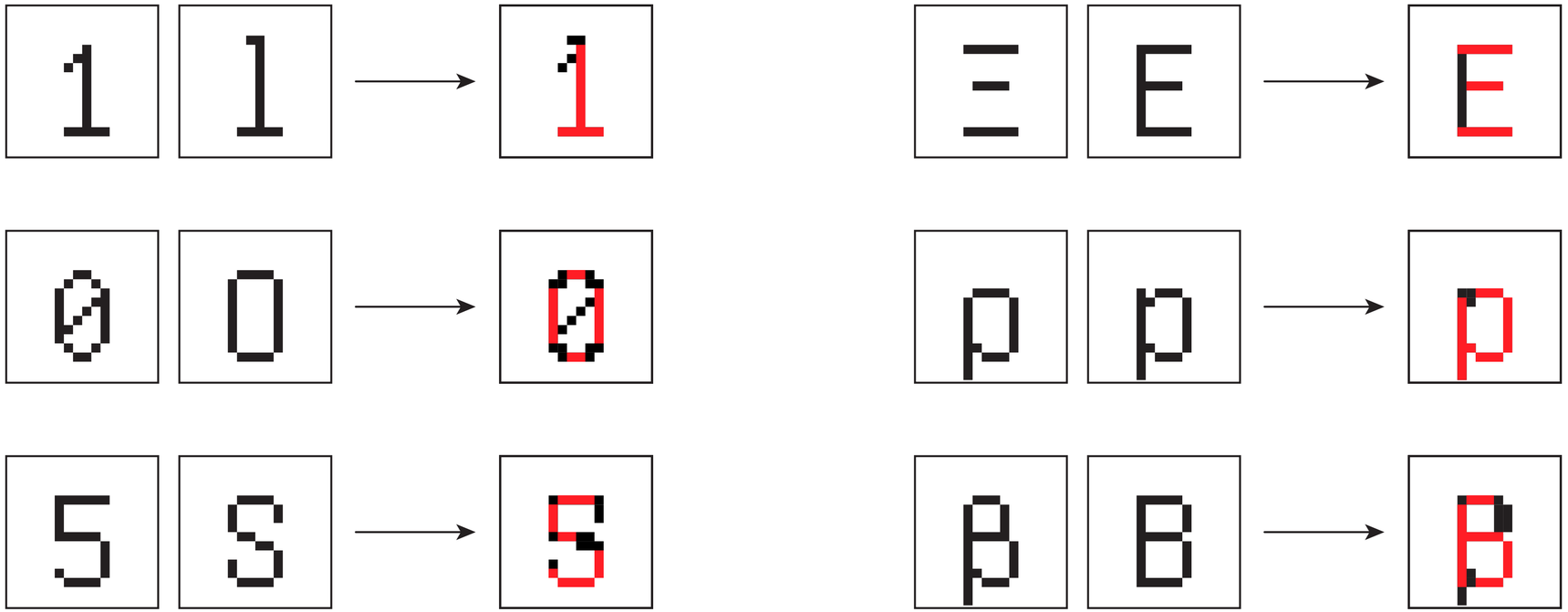
Visual Archetypes Transformer

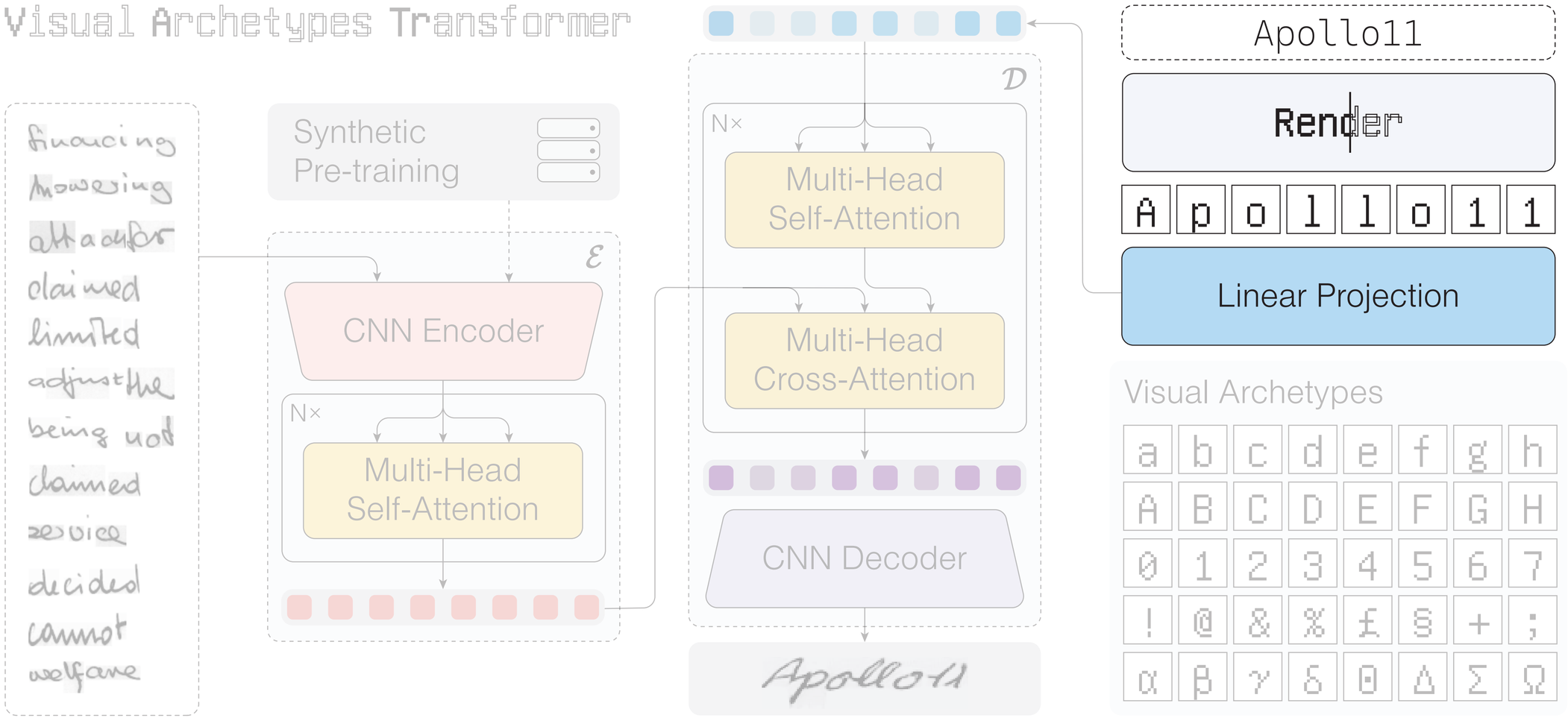


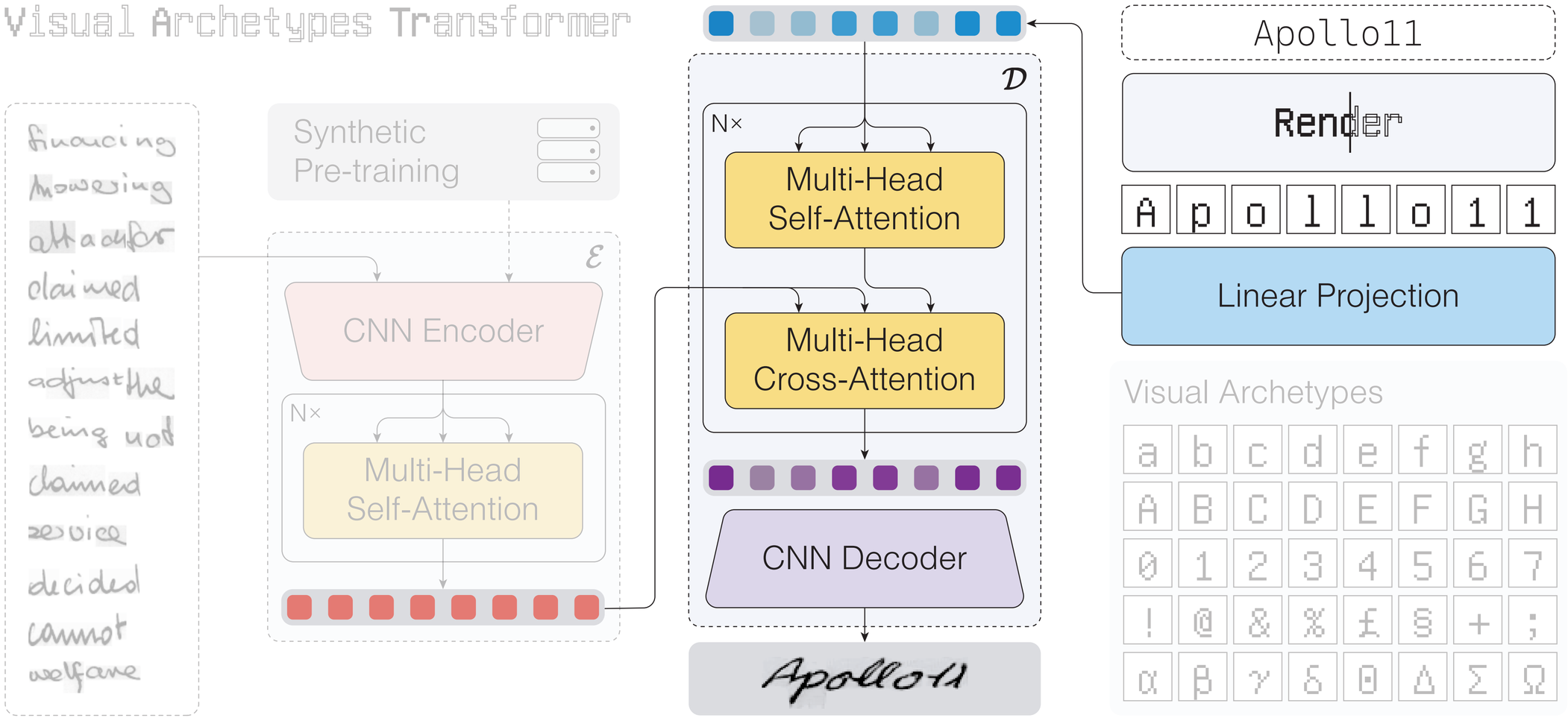

Style examples
Apollo11
One-hot vectors
HWT


Style examples
Apollo11
One-hot vectors
Visual archetypes
A p o l
l o 1 1
HWT
VATr

Style	Content	HWT	VATr
	Beryl		
	500		
	2		
	Federal		



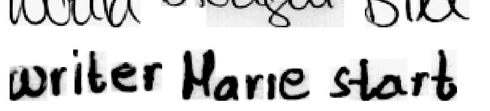

Visual Archetypes Transformer
Synthetic
Pre-training
CNN Encoder
N×
Multi-Head
Self-Attention
Multi-Head
Self-Attention
Multi-Head
Cross-Attention
CNN Decoder
Apollo11
Render
A p o l l o 1 1
Linear Projection
Visual Archetypes
a b c d e f g h
A B C D E F G H
0 1 2 3 4 5 6 7
! @ & % £ § + ;
α β γ δ θ Δ Σ Ω


Visual Archetypes Transformer
financing
knowing
attacker
claimed
limited
adjust the
being not
claimed
service
decided
cannot
welfare
Synthetic
Pre-training
$\mathcal{E}$
CNN Encoder
N×
Multi-Head
Self-Attention
$\mathcal{D}$
N×
Multi-Head
Self-Attention
Multi-Head
Cross-Attention
CNN Decoder
Apollo11
Apollo11
Render
A p o l l o 1 1
Linear Projection
Visual Archetypes
a b c d e f g h
A B C D E F G H
0 1 2 3 4 5 6 7
! @ & % £ § + ;
α β γ δ θ Δ Σ Ω

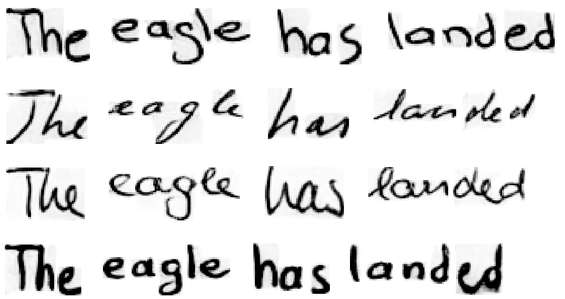
	ALL	IV-S	IV-U	OOV-S	OOV-U
TS-GAN	20.65	118.56	128.75	127.11	136.67
GANwriting	20.55	120.07	124.30	125.87	130.68
HWT	19.40	106.97	108.84	109.45	114.10
VATr (Ours)	**17.79**	**88.20**	**91.11**	**98.57**	**102.22**

	Long-Tail		Digits	
	FID	CER	FID	CER
HiGAN	26.08	**8.63**	26.08	101.53
HWT	40.95	20.36	40.95	98.47
VATr (Ours)	**21.36**	11.85	**21.36**	**94.66**

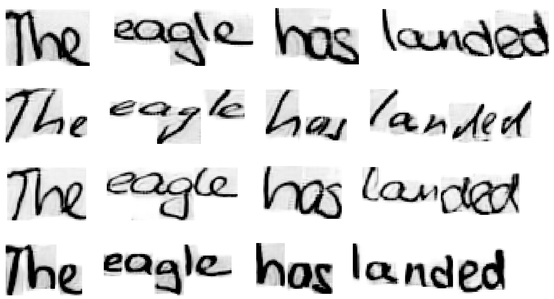
Style examples



VATr



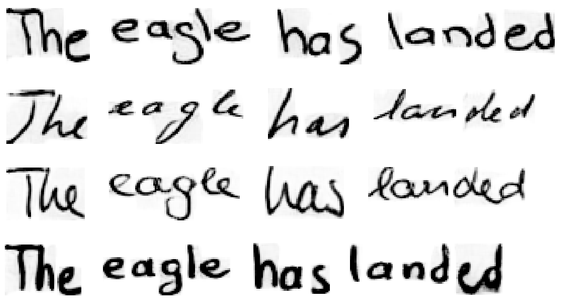
HWT



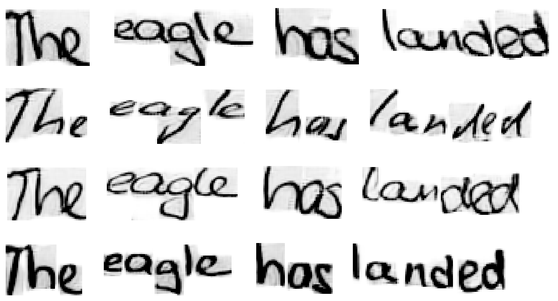
- The pretrained **style encoder** stabilizes the training process

Style examples

VATr



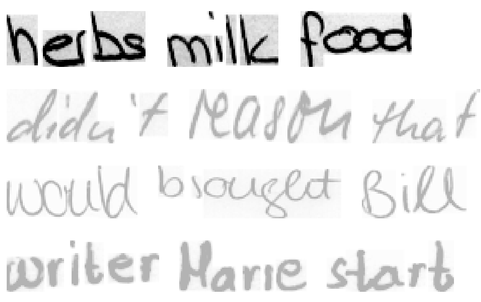
HWT



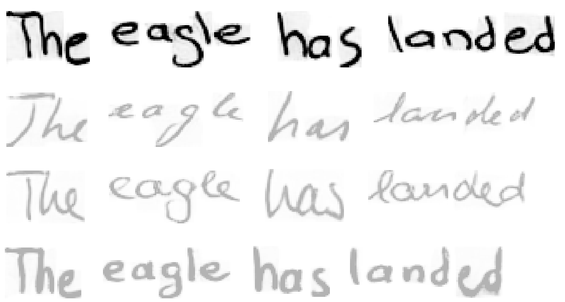
- The pretrained **style encoder** stabilizes the training process

- The pretrained **style encoder** reduces the artifacts in the output images.

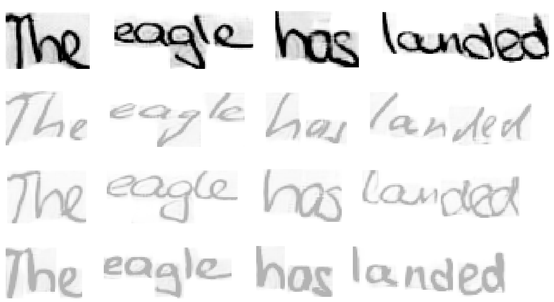
Style examples



VATr



HWT



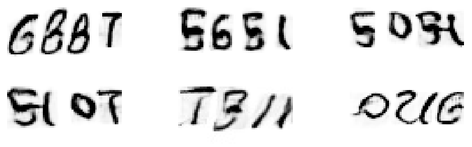
- The pretrained **style encoder** stabilizes the training process
- The pretrained **style encoder** reduces the artifacts in the output images.
- The **Visual Archetypes** better exploit the repetitive visual patterns of the characters.

Content

6887 5651 5051

5107 7311 0216

VATr

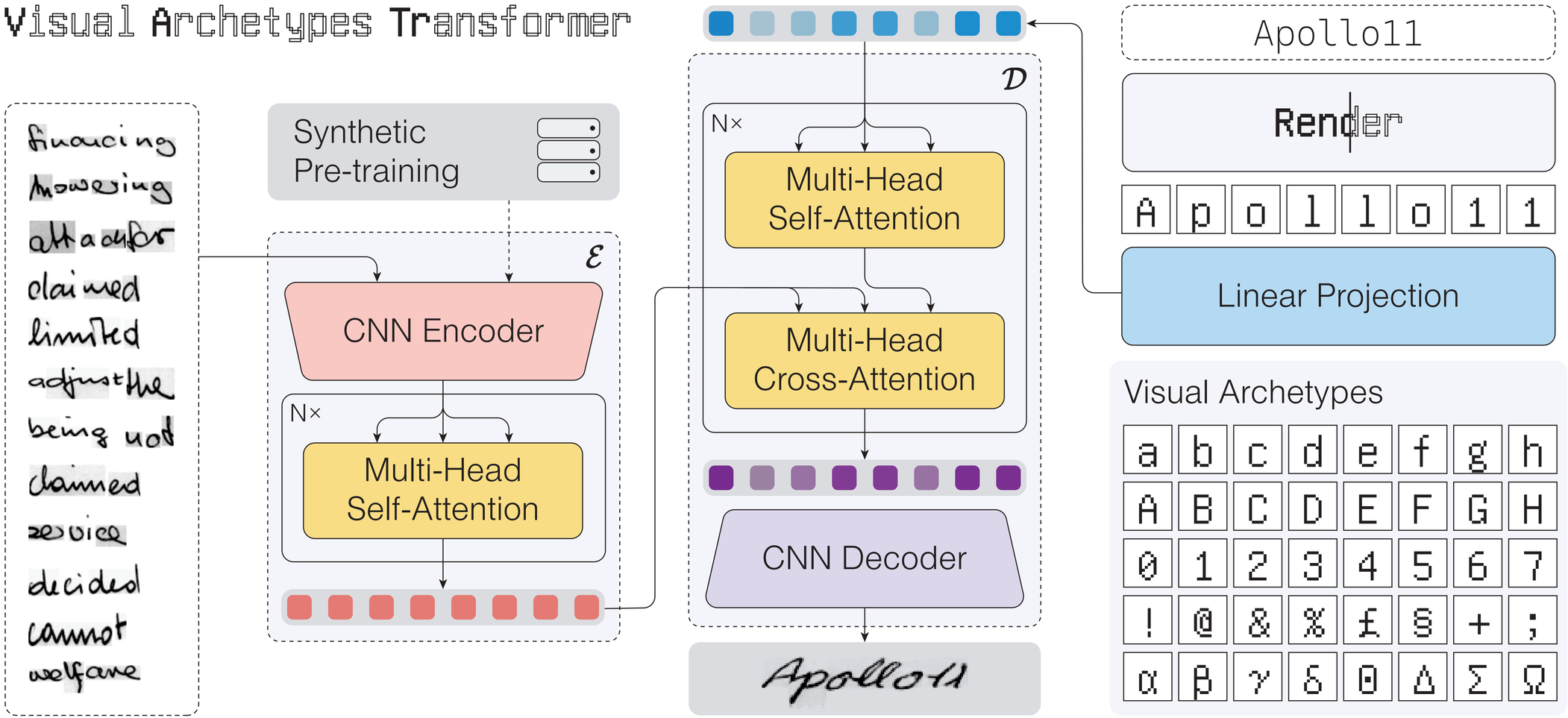


HWT

- The pretrained **style encoder** stabilizes the training process
- The pretrained **style encoder** reduces the artifacts in the output images.
- The **Visual Archetypes** better exploit the repetitive visual patterns of the characters.
- The **Visual Archetypes** allow handling out-of-charset characters.

Style examples

α δ η μ ν ρ Σ ω


Visual Archetypes Transformer
financing
showing
attacker
claimed
limited
adjust the
being not
claimed
service
decided
cannot
welfare
Synthetic
Pre-training
$\mathcal{E}$
CNN Encoder
N×
Multi-Head
Self-Attention
$\mathcal{D}$
N×
Multi-Head
Self-Attention
Multi-Head
Cross-Attention
CNN Decoder
Apollo11
Apollo11
Render
A p o l l o 1 1
Linear Projection
Visual Archetypes
a b c d e f g h
A B C D E F G H
0 1 2 3 4 5 6 7
! @ & % £ § + ;
α β γ δ θ Δ Σ Ω

AImage Lab

UNIMORE
UNIVERSITÀ DEGLI STUDI DI
MODENA E REGGIO EMILIA

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Thanks for your attention!

Vittorio Pippi Silvia Cascianelli Rita Cucchiara

{name}.{surname}@unimore.it