Style Image Retrieval for Improving Material Translation Using Neural Style Transfer

***Gibran Benitez-Garcia*[1,2]**, Wataru Shimoda, and Keiji Yanai[2]

OMRON SINIC X Corporation[1],
The University of Electro-Communications[2]

MMArt-ACM '20 2020.10.26

Motivation

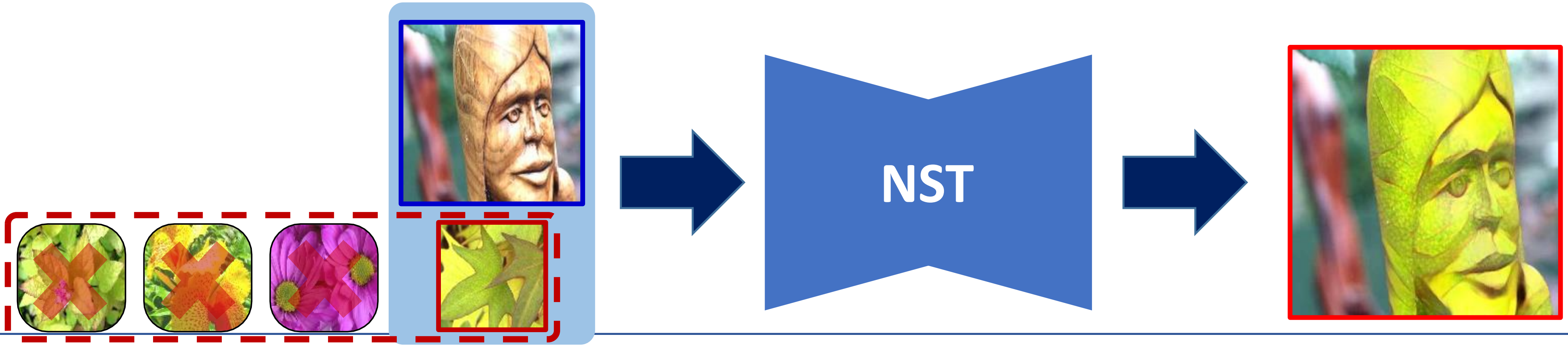
– *Results of Neural Style Transfer methods to* ***translate object materials*** *rely on the style picture chosen to modify the content image*

– ***Automatically find the ideal style image*** *that better translates the material of an object*

Content Image **Translated images with different materials**

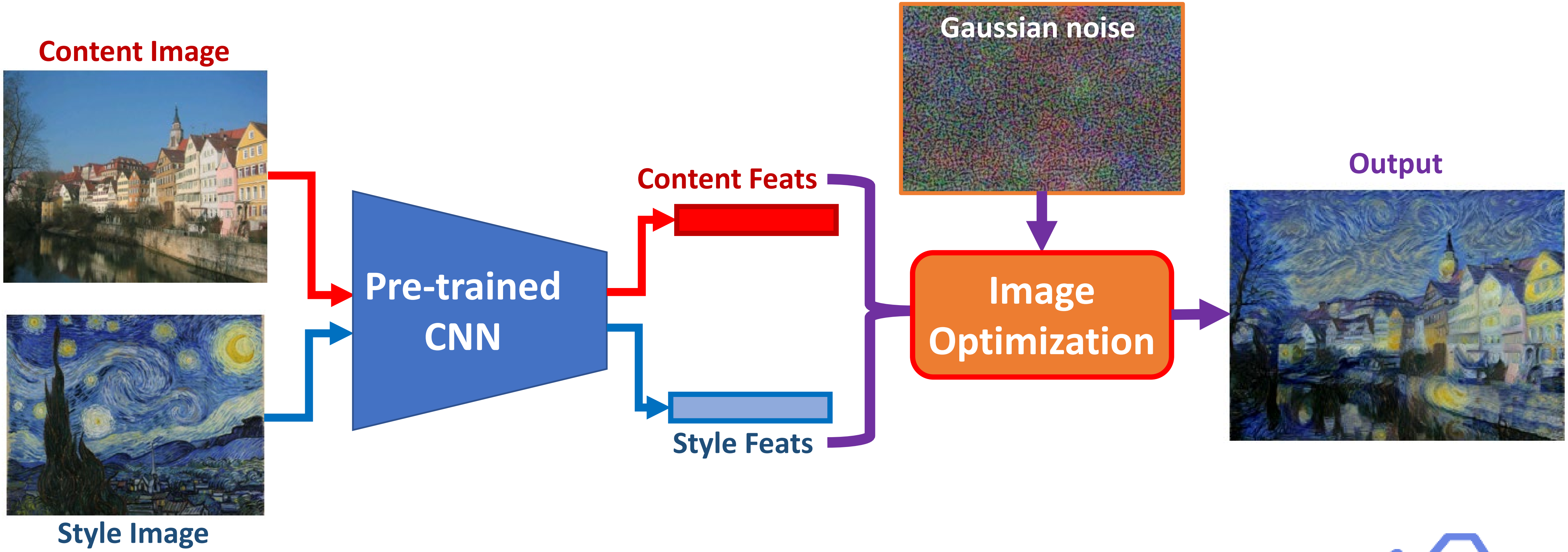
Objective

- ***Objective:***
  - *Automatically find the ideal style image based on its discrimination level and its relation with the content image in terms of semantic information*
- ***Approach:***
  - *An image retrieval method based on the most discriminative candidate style images, and evaluate the semantic similarity with the content using IN*



Neural Style Transfer

– *NST* exploits **CNN feature activations** to recombine the **content** of a given photo and the **style** of artworks*

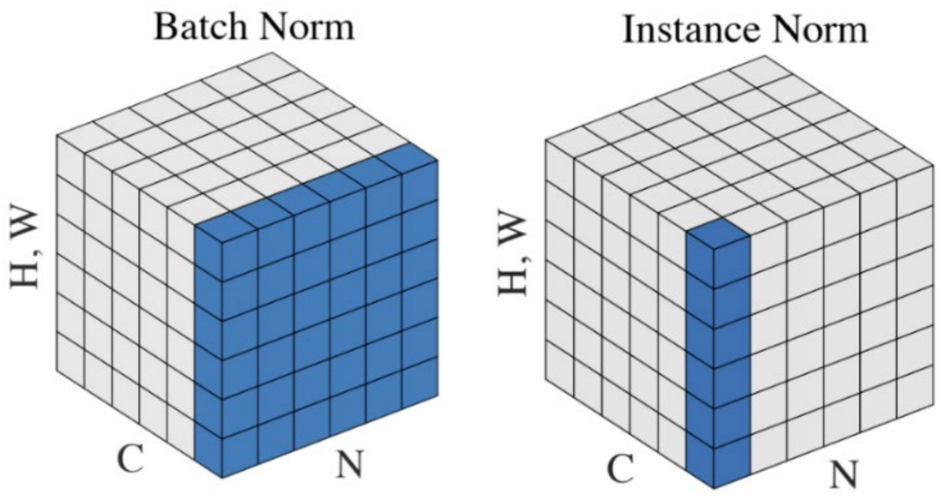


 * Gatys, Leon, et al. "Image style transfer using convolutional neural networks." CVPR 2016.

Instance Normalization

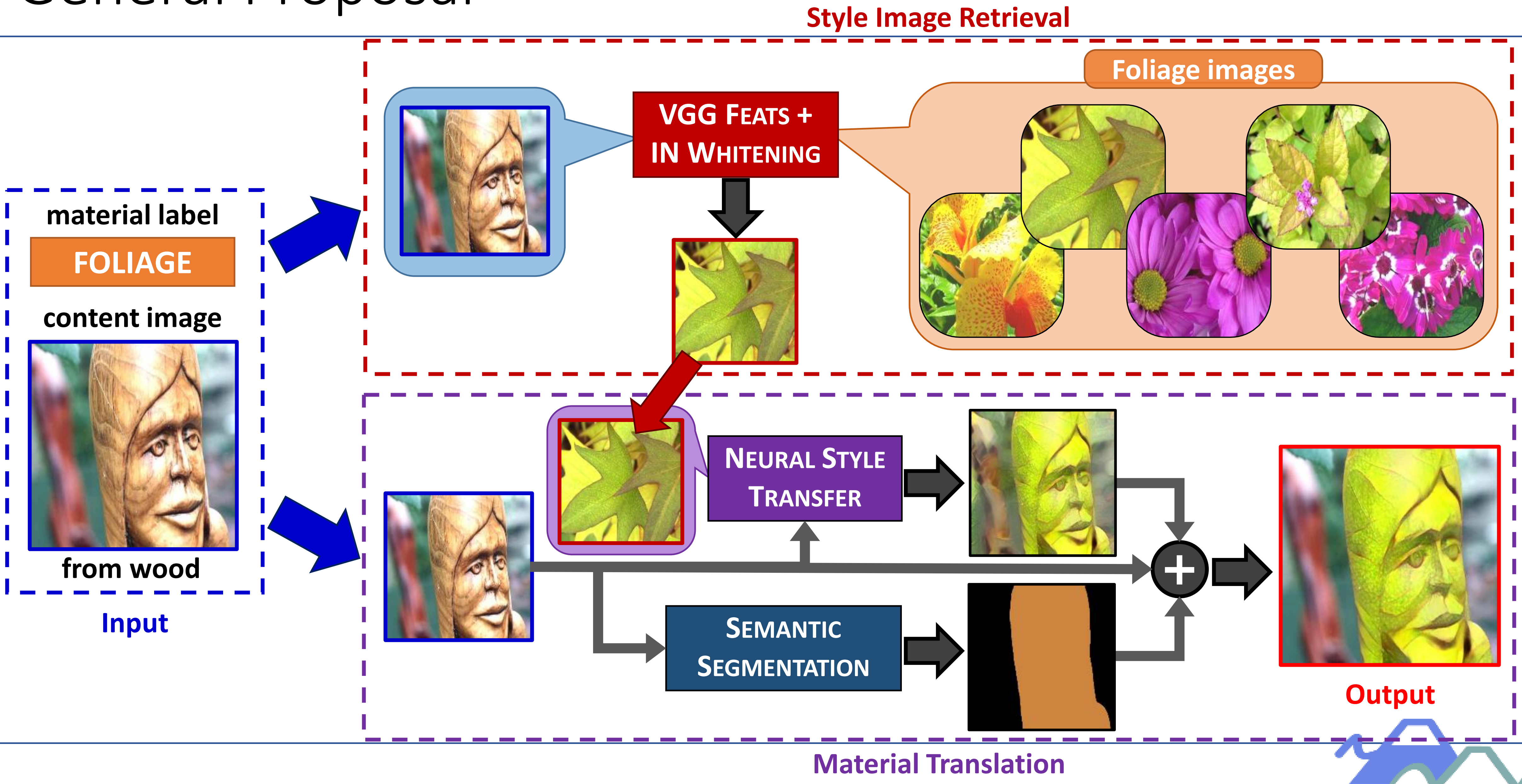
– *IN[*] computes the mean/standard deviation and* ***normalize across each channel*** *in each training example*

– *Generates a network agnostic to the contrast of the original images. A.k.a.* ***erases style information****.*



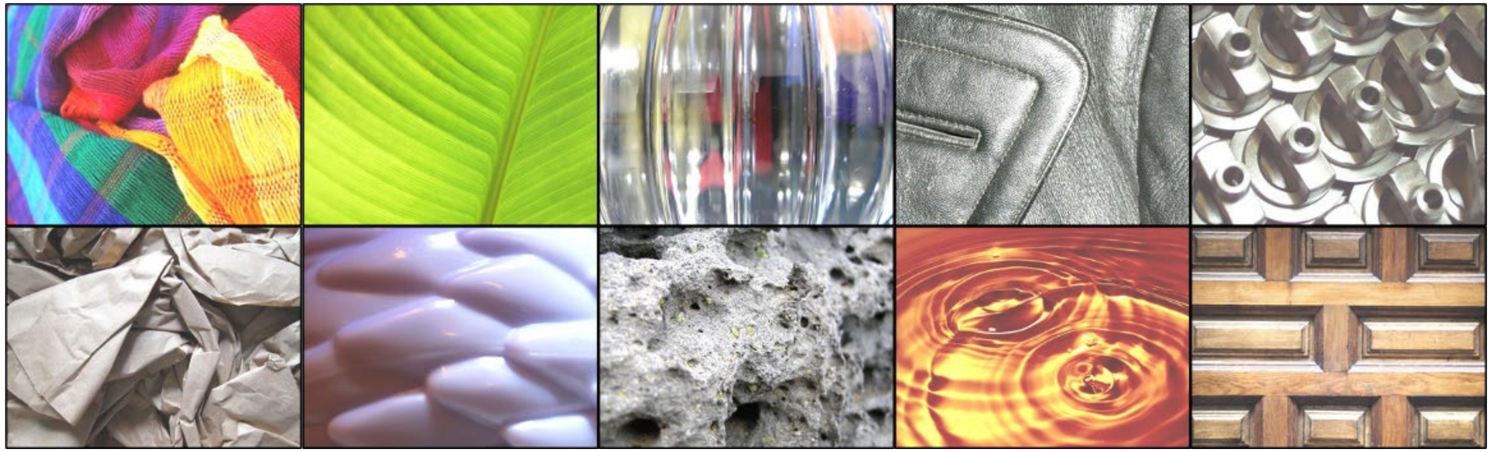
* Ulyanov, Dmitry, et al. "Instance normalization: The missing ingredient for fast stylization." arXiv 2016.

General Proposal



Style Image Retrieval

1) ***Search refinement***
   – *Choose the **best scored material** image per class (pre-trained CNN), and the images with more **extensive material regions** (segmentation)*
2) ***Style removal***
   – ***IN whitening*** *from the pre-trained CNN, and* ***L2 norm*** *from query and styles*



Datasets of object materials

- ***FMD* dataset:***
  - *Pixel labels*
  - *10 class materials*
  - *1,000 images in total*

- ***Extended-FMD** dataset:***
  - *Image labels*
  - *Same classes as FMD*
  - *10,000 images in total*

* Sharan, Lavanya, et al. "Material perception: What can you see in a brief glance?." Journal of Vision, 2009.

** Zhang, Yan, et al. "Integrating deep features for material recognition." ICPR, 2016.

Evaluation of IN-based style retrieval

- ***Classification and segmentation metrics*** *to evaluate generated results: average accuracy (acc) and mean Intersection over the Union (mIoU).*
- *Baseline: fixed style images, all processes based on* ***Gatys NST***

	w/o refine		w/ refine	
Method	acc	mIoU	acc	mIoU
Baseline	-	-	0.556	0.4860
VGG19-IN	**0.409**	**0.3967**	**0.572**	**0.5062**
VGG19-BN	0.291	0.3612	0.543	0.4887
VGG19	0.270	0.3520	0.506	0.4845

Cross-classification of material translation

original materials (material A) \ translated materials (material B)	Fa	Fo	Gl	Le	Me	Pa	Pl	St	Wa	Wo
Fabric	-	64	88	27	68	62	80	72	23	62
Foliage	23	-	70	11	27	24	38	40	12	50
Glass	47	38	-	20	55	41	71	41	22	63
Leather	86	32	81	-	35	21	63	54	**6**	85
Metal	69	27	**94**	37	-	28	56	62	10	80
Paper	47	24	32	16	27	-	65	49	11	52
Plastic	68	33	86	45	73	26	-	72	30	48
Stone	71	66	87	**7**	49	72	74	-	23	**94**
Water	36	27	68	**4**	46	37	41	58	-	74
Wood	48	52	90	39	33	33	**97**	84	**9**	-

Qualitative results

– ***Wood** object to different materials, using NST and IN-based style retrieval*

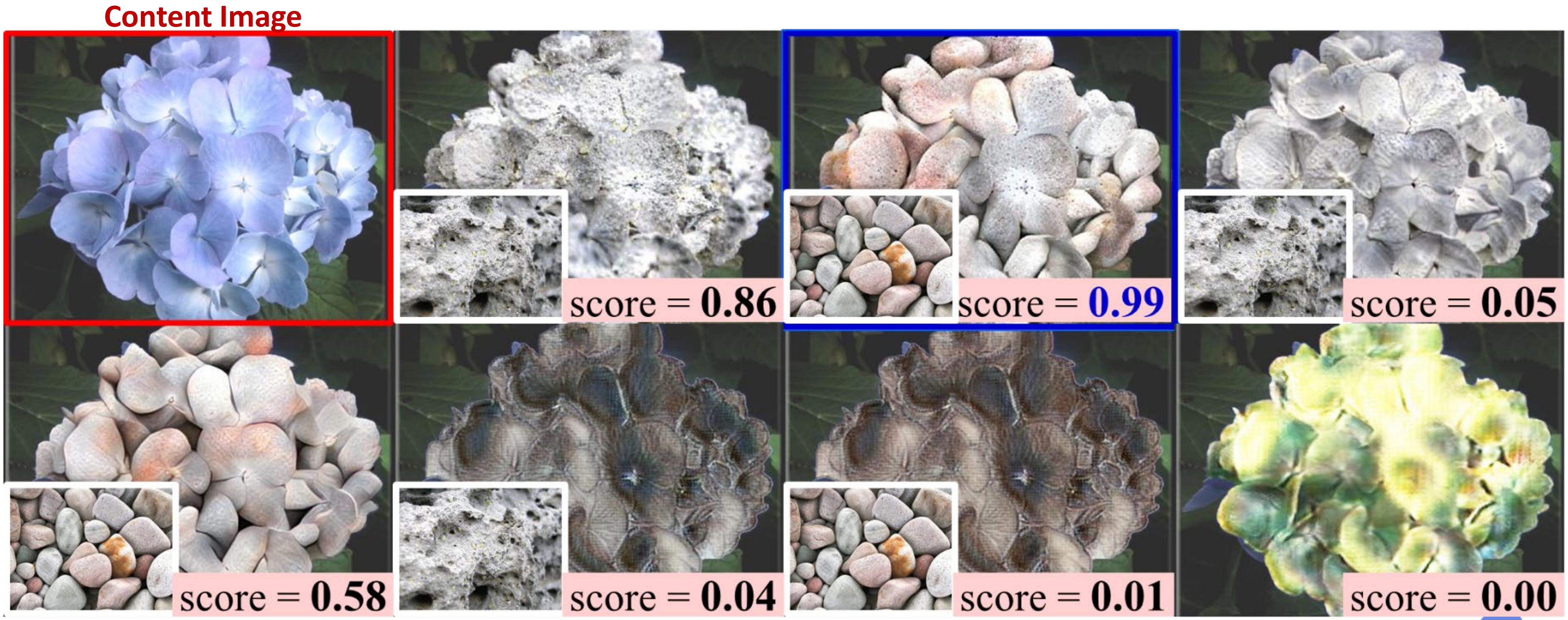
Comparison with different NST methods

– *We evaluated all methods using* ***GAN metrics****, i.e., Inception Score (IS), and the Frechet Inception Distance (FID).*

Method	Acc ↑	mIoU ↑	IS ↑	FID ↓
NST-Base	0.556	0.4860	4.161	66.54
NST-IN (ours)	**0.572**	**0.5062**	**4.181**	**61.30**
WCT-Base	0.349	0.4133	3.518	65.61
WCT-IN	0.353	0.4079	3.604	64.53
MUNIT-Base	0.343	0.3872	3.475	65.60
MUNIT-IN	0.373	0.3995	3.523	61.52
StarGAN	0.113	0.2738	2.673	103.8

Qualitative comparison

*NST-Base, **NST-IN**, WCT-Base, WCT-IN, MUNIT-Base, MUNIT-IN, and StarGAN*



Conclusion

- ***Conclusions:***
  - *We experimentally proved that by defining an image style search with IN, the results of NST material translation are significantly better.*
  - *Our style retrieval proposal can boost material translation results of conventional NST methods, such as Gatys, WCT, and MUNIT.*
- ***Future work:***
  - *Test and analyze different options for removing the style information.*
  - *Integrate the NST, Segmentation, and Search process for results optimization.*

Thank You