Réseaux de Neurones Légers Multi-Echelle pour la restauration stylisée et contrôlée d'images - journée normastic - 8 avril 2022

{**Thibault.Durand**, Julien.Rabin & David.Tschumperlé} @unicaen.fr

Normandie Université
Equipe Image
GREYC UMR 6072
14000 Caen, France

Problématique - restauration d'images

Problématique - restauration d'images

- X (resp. x) $\in \mathbb{R}^{K \times N \times N \times 3}$ est un tenseur de K patchs originaux (resp. patchs dégradés) de taille $N \times N$

Problématique - restauration d'images

- X (resp. x) $\in \mathbb{R}^{K \times N \times N \times 3}$ est un tenseur de K patchs originaux (resp. patchs dégradés) de taille $N \times N$
- $x_k \in \mathbb{R}^{N \times N \times 3}$ et $X_k \in \mathbb{R}^{N \times N \times 3}$ sont des élements de X et x.

Problématique - restauration d'images

- X (resp. x) $\in \mathbb{R}^{K \times N \times N \times 3}$ est un tenseur de K patchs originaux (resp. patchs dégradés) de taille $N \times N$
- $x_k \in \mathbb{R}^{N \times N \times 3}$ et $X_k \in \mathbb{R}^{N \times N \times 3}$ sont des élements de X et x.
- D est l'opérateur de dégradation tel que $D(X) = x$.

Problématique - restauration d'images

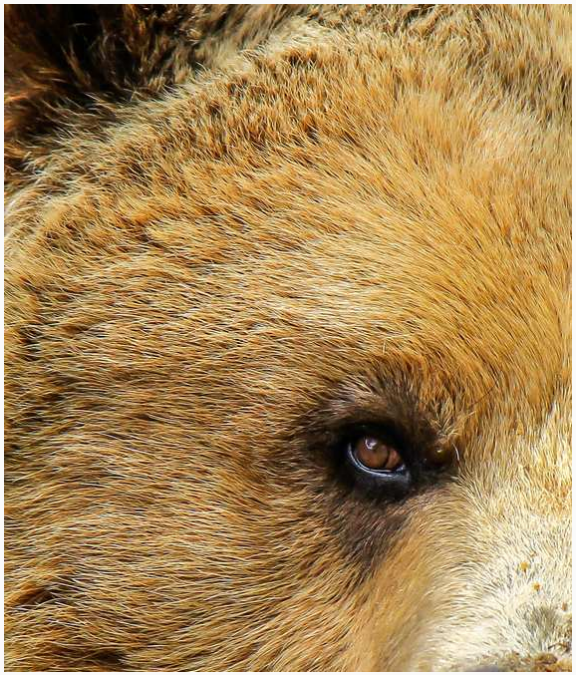
- X (resp. x) $\in \mathbb{R}^{K \times N \times N \times 3}$ est un tenseur de K patchs originaux (resp. patchs dégradés) de taille $N \times N$
- $x_k \in \mathbb{R}^{N \times N \times 3}$ et $X_k \in \mathbb{R}^{N \times N \times 3}$ sont des élements de X et x.
- D est l'opérateur de dégradation tel que $D(X) = x$.
- $\hat{X}$ correspond à la prédiction du réseau à partir de x

Problématique - restauration d'images

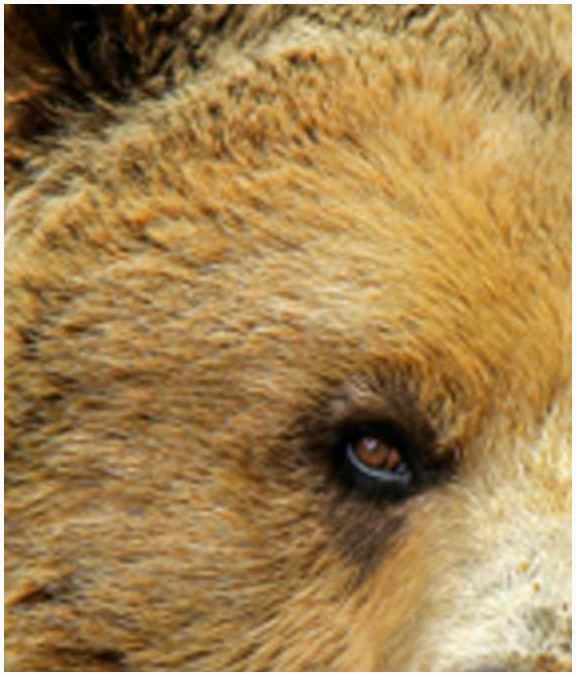
- X (resp. x) $\in \mathbb{R}^{K \times N \times N \times 3}$ est un tenseur de K patchs originaux (resp. patchs dégradés) de taille $N \times N$
- $x_k \in \mathbb{R}^{N \times N \times 3}$ et $X_k \in \mathbb{R}^{N \times N \times 3}$ sont des élements de X et x.
- D est l'opérateur de dégradation tel que $D(X) = x$.
- $\hat{X}$ correspond à la prédiction du réseau à partir de x

$\rightarrow$ Quel est l'objectif de la restauration ? Est ce de retrouver X à partir de x?

Problématique - Exemple de la Super-Résolution (x4)

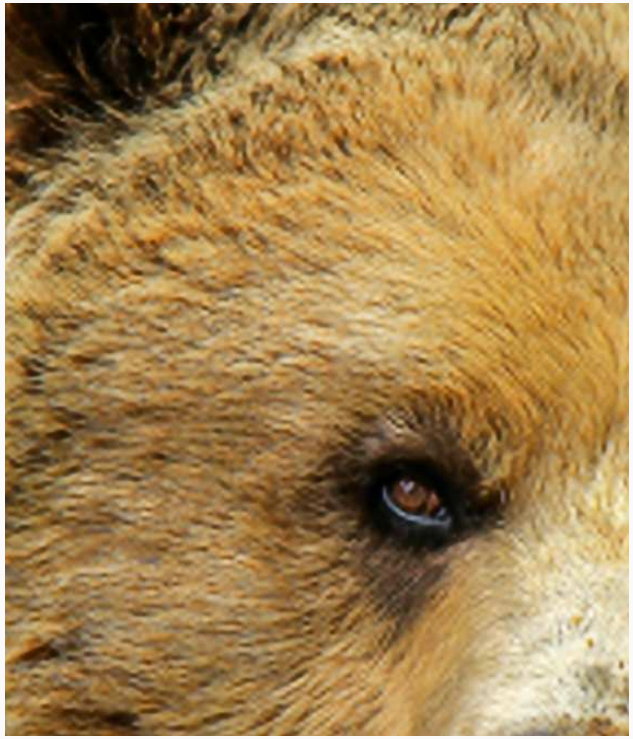


X_k



x_k

Problématique - Exemple de la Super-Résolution (x4)



Premiers réseaux convolutionnels

SRCNN[1]
(entre $\sim$ 100k et $\sim$ 400k paramètres)

$$\mathcal{L}_{\text{MSE}}(X, \hat{X}) = \frac{1}{K} \sum_{k=1}^{K} \|X_k - \hat{X}_k\|^2$$

[1] Chao Dong et al. (2014). "Learning a Deep Convolutional Network for Image Super-Resolution". In: *Computer Vision – ECCV 2014*. Ed. by David Fleet et al. Cham: Springer International Publishing, pp. 184–199. ISBN: 978-3-319-10593-2.

Problématique - Exemple de la Super-Résolution (x4)

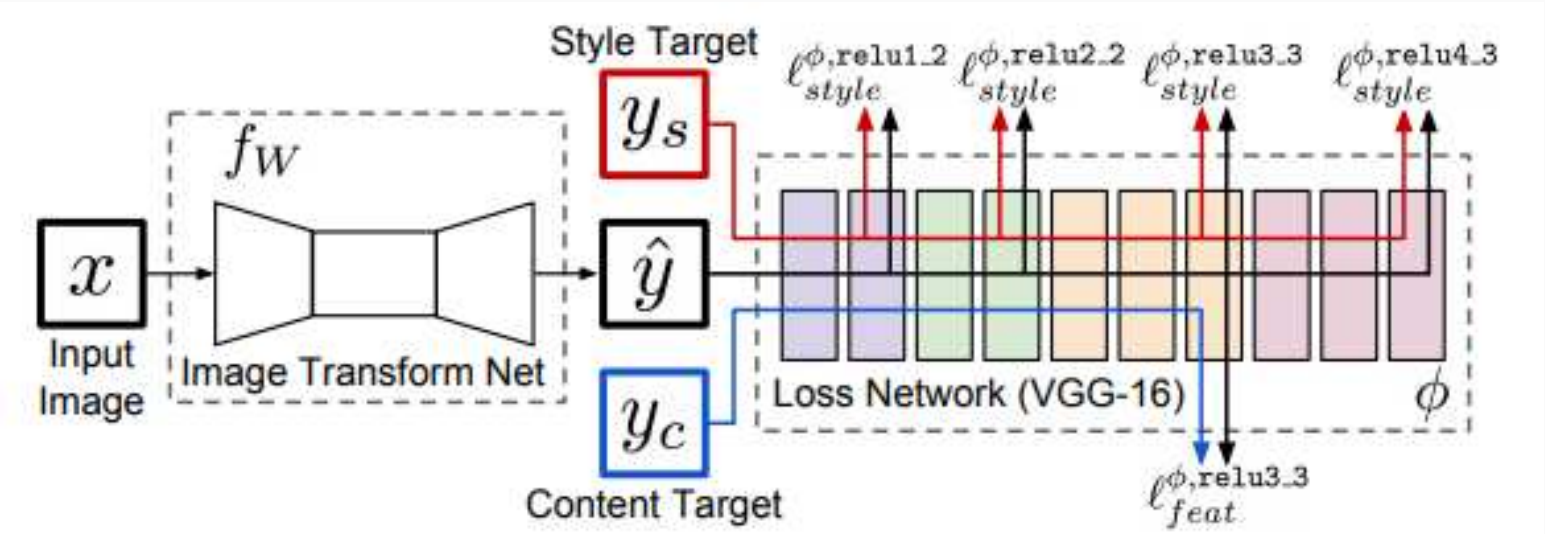
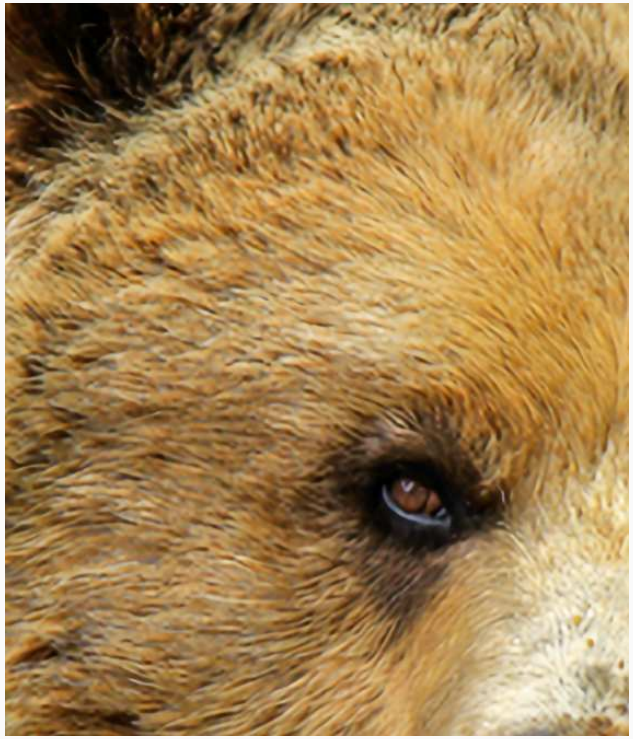


Figure 1: Fonctionnement du réseau VGG-16 pour l'extraction de caractéristiques profondes et abstraites, dîtes ' de style'. Source de la figure : [1].

[1] Justin Johnson, Alexandre Alahi, and Li Fei-Fei (Oct. 2016). "Perceptual Losses for Real-Time Style Transfer and Super-Resolution". In: vol. 9906, pp. 694–711. ISBN: 978-3-319-46474-9. DOI: 10.1007/978-3-319-46475-6_43

Problématique - Exemple de la Super-Résolution (x4)



EDSR[2]
(∼1500K paramètres)

$$\mathcal{L}_{\text{MSE}}(X, \hat{X}) = \frac{1}{K} \sum_{k=1}^{K} \|X_k - \hat{X}_k\|^2$$

$$\mathcal{L}_{\text{Perc}}(X, \hat{X}) = \frac{1}{K} \sum_{\ell \in L_{\text{Perc}}} \|\phi_\ell(X) - \phi_\ell(\hat{X})\|^2$$

où $\phi_\ell(.)$ correspond aux *réponses* normalisées de la ℓ-ième couche d'un classifieur préentrainé

[2] Bee Lim et al. (2017). "Enhanced Deep Residual Networks for Single Image Super-Resolution". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1132–1140. DOI: 10.1109/CVPRW.2017.151.

Problématique - Exemple de la Super-Résolution

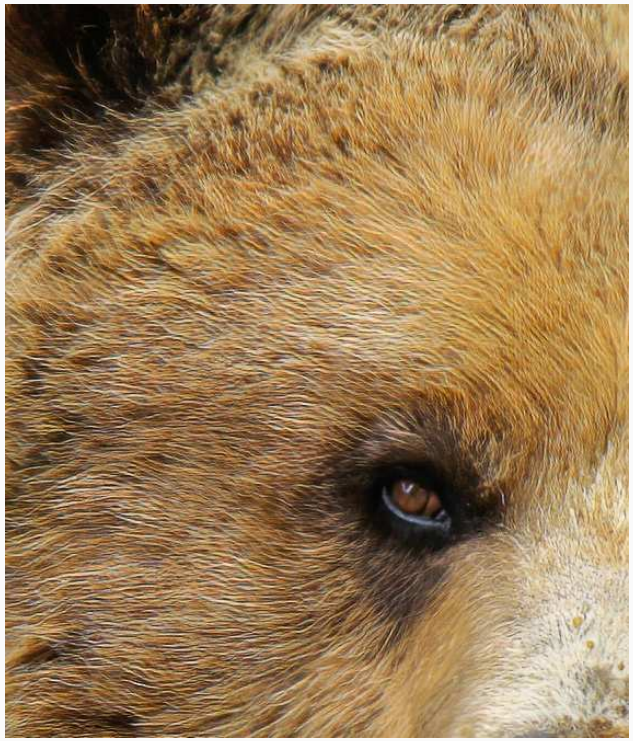
X_k	x_k	$\hat{X}^3$	$\hat{X}^4$
PSNR / SSIM	22.75 / 0.5946	23.42 / 0.6168	21.90 / 0.6083

Figure 2: Illustration, dans le cadre du problème de Super-Résolution, de non corrélation entre le PSNR et l'évaluation subjective. Source : [4]

[3] Chao Dong et al. (2014). "Learning a Deep Convolutional Network for Image Super-Resolution". In: *Computer Vision – ECCV 2014*. Ed. by David Fleet et al. Cham: Springer International Publishing, pp. 184–199. ISBN: 978-3-319-10593-2.

[4] Justin Johnson, Alexandre Alahi, and Li Fei-Fei (Oct. 2016). "Perceptual Losses for Real-Time Style Transfer and Super-Resolution". In: vol. 9906, pp. 694–711. ISBN: 978-3-319-46474-9. DOI: 10.1007/978-3-319-46475-6_43.

Problématique - Exemple de la Super-Résolution (x4)



ESRGAN[5]
(∼1500K paramètres)

Pénalités adverses ;
réseaux antagonistes

[5] Xintao Wang et al. (Sept. 2018). *ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks*.

Motivations

- Le problème de SR est un problème double. La restauration des structures, des contours, mais aussi la restauration des textures.
  $\rightarrow$ *Nous choisissons de résoudre le problème en deux temps*

Motivations

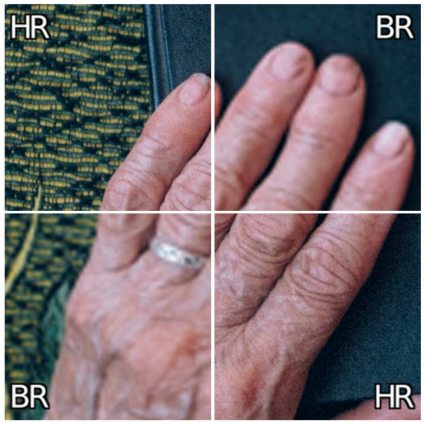
- Le problème de SR est un problème double. La restauration des structures, des contours, mais aussi la restauration des textures.
  $\rightarrow$ *Nous choisissons de résoudre le problème en deux temps*
- Les méthodes proposées sont souvent de plus en plus lourdes en nombre de paramètres.
  $\rightarrow$ *Nous nous limitons à des réseaux légers* $\sim$ *Mo*

Motivations

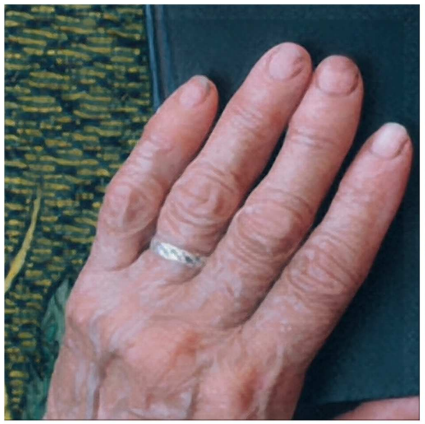
- Le problème de SR est un problème double. La restauration des structures, des contours, mais aussi la restauration des textures.
  $\rightarrow$ *Nous choisissons de résoudre le problème en deux temps*
- Les méthodes proposées sont souvent de plus en plus lourdes en nombre de paramètres.
  $\rightarrow$ *Nous nous limitons à des réseaux légers* $\sim$ *Mo*
- Le problème de SR n'est, à priori, jamais envisagé comme un problème inverse mal posé, c'est à dire comme un problème où la solution peut (doit?) dépendre de contraintes en entrée, et où plusieurs solutions sont possibles.
  $\rightarrow$ *Nous proposons du contrôle à l'utilisateur quant au choix des textures à insérer*

Contributions - Présentation

Réseaux de neurones légers pour la restauration stylisée d'images (Ici, Super-Résolution Stylisée)



HR & BR (interp. bicubique)


Résultat de l'étape-SR ($\times 4$)


Résultat de l'étape-ST

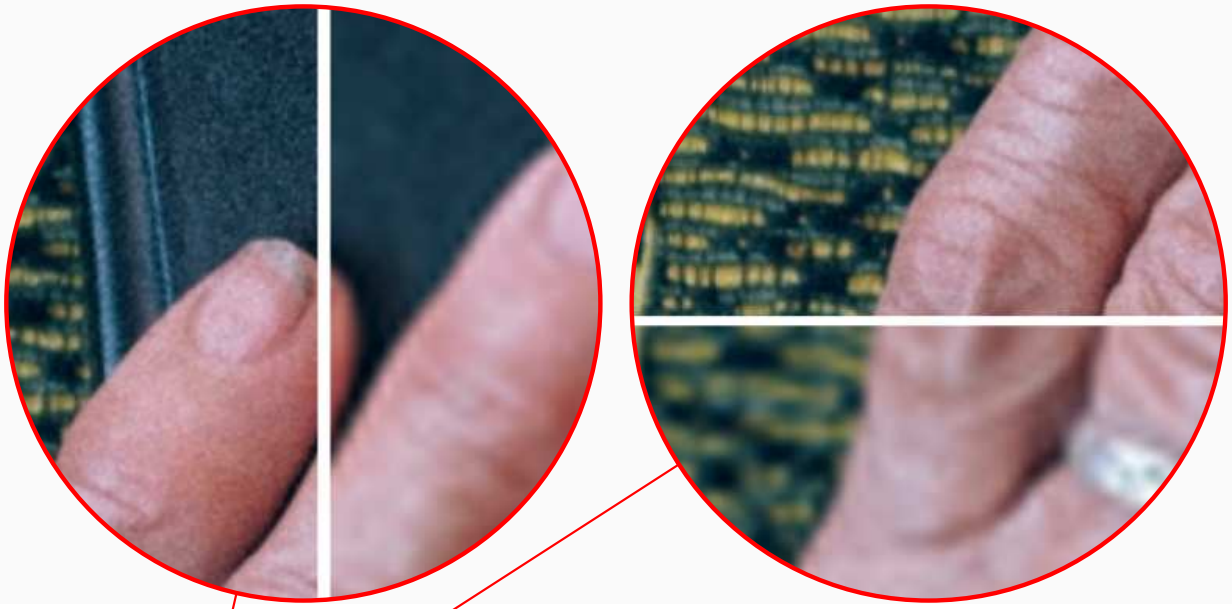

Masques β_i

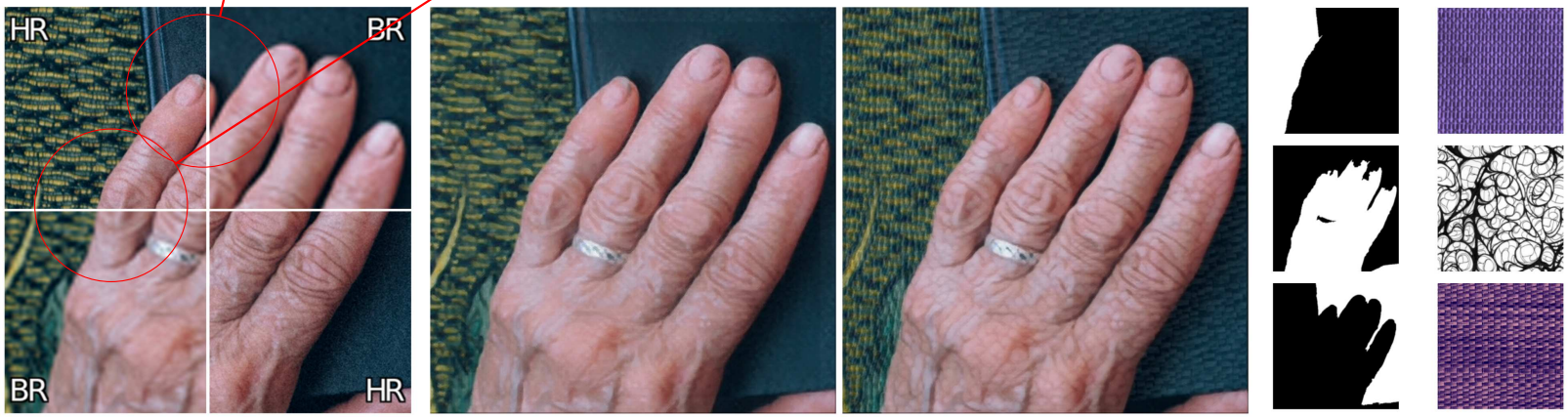

Styles Y_i

Une approche en **deux étapes** :

- Etape de Restauration (ici, étape-SR) ($\sim$ 200k paramètres ; ordre de grandeur **xMo**).
  Utilisation d'un réseau **multi-échelle** pour une bonne **reconstruction des contours, de la géométrie, sans contrôle**.

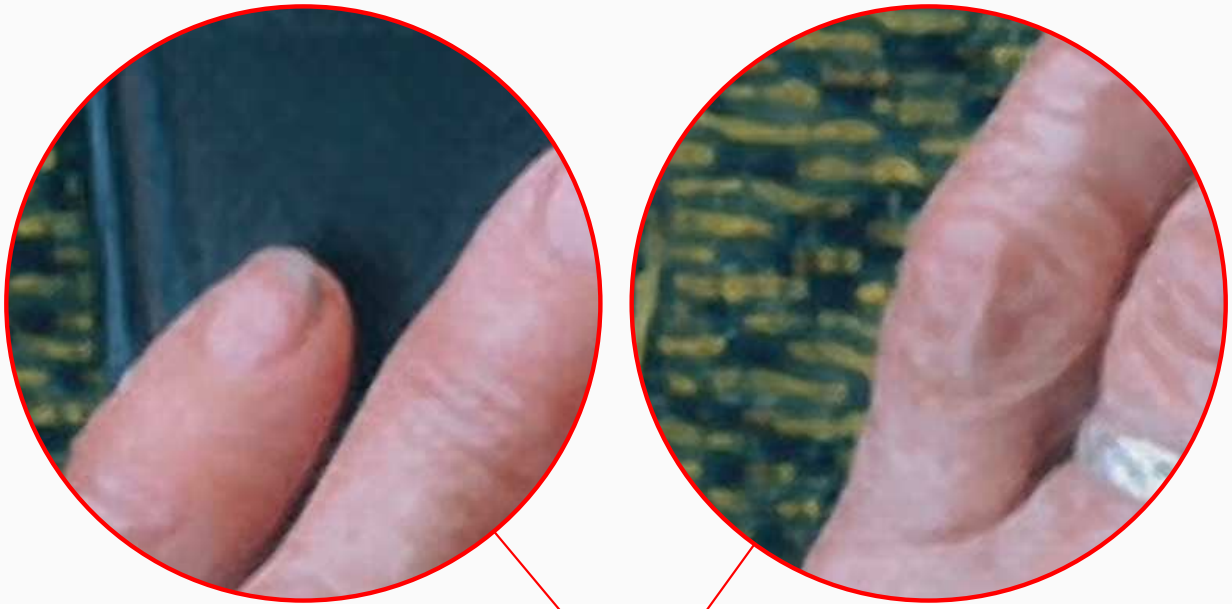
Contributions - Présentation



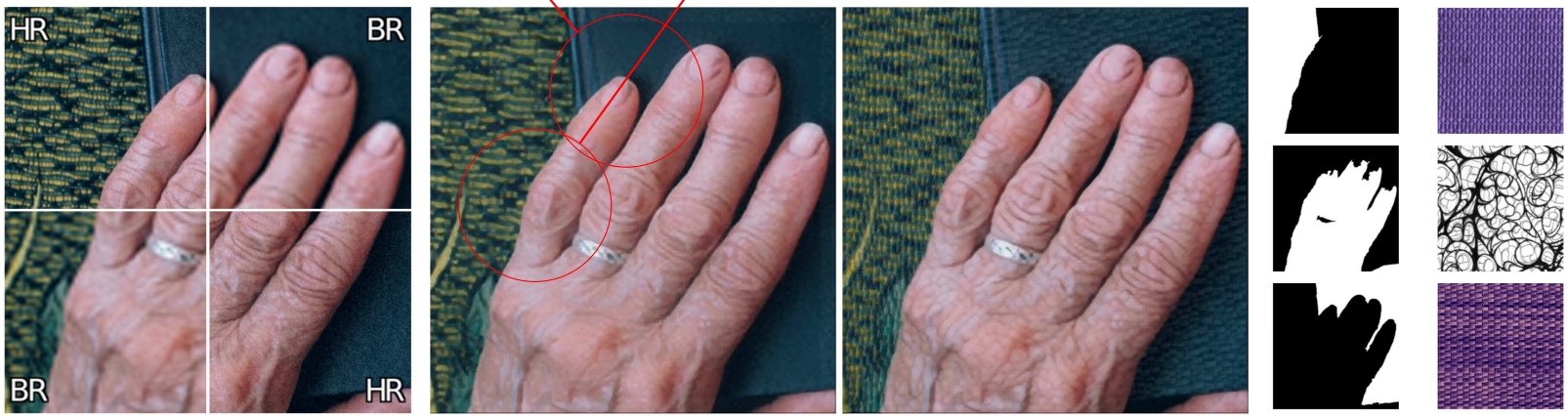


HR & BR (interp. bicubique) | Résultat de l'étape-SR (×4) | Résultat de l'étape-ST | Masques β_i | Styles Y_i

Contributions - Présentation



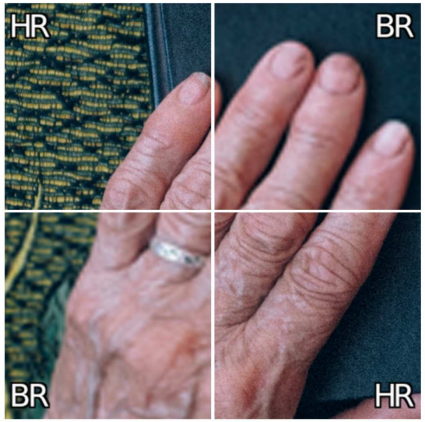
Etape 1



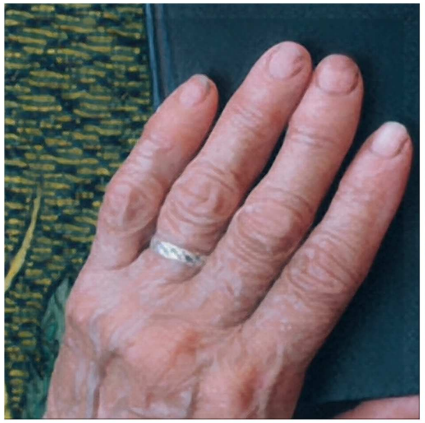
HR & BR (interp. bicubique) | Résultat de l'étape-SR (×4) | Résultat de l'étape-ST | Masques β_i | Styles Y_i

Contributions - Présentation

Réseaux de neurones légers pour la restauration stylisée d'images (Ici, Super-Résolution Stylisée)



HR & BR (interp. bicubique)


Résultat de l'étape-SR ($\times 4$)


Résultat de l'étape-ST

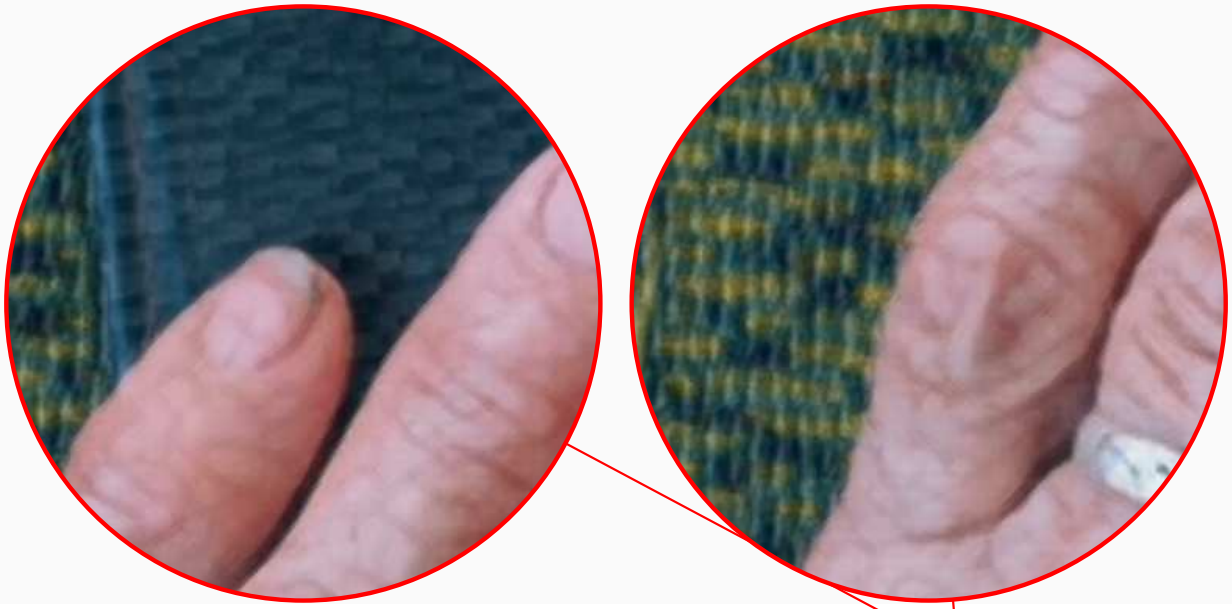

Masques β_i


Styles Y_i

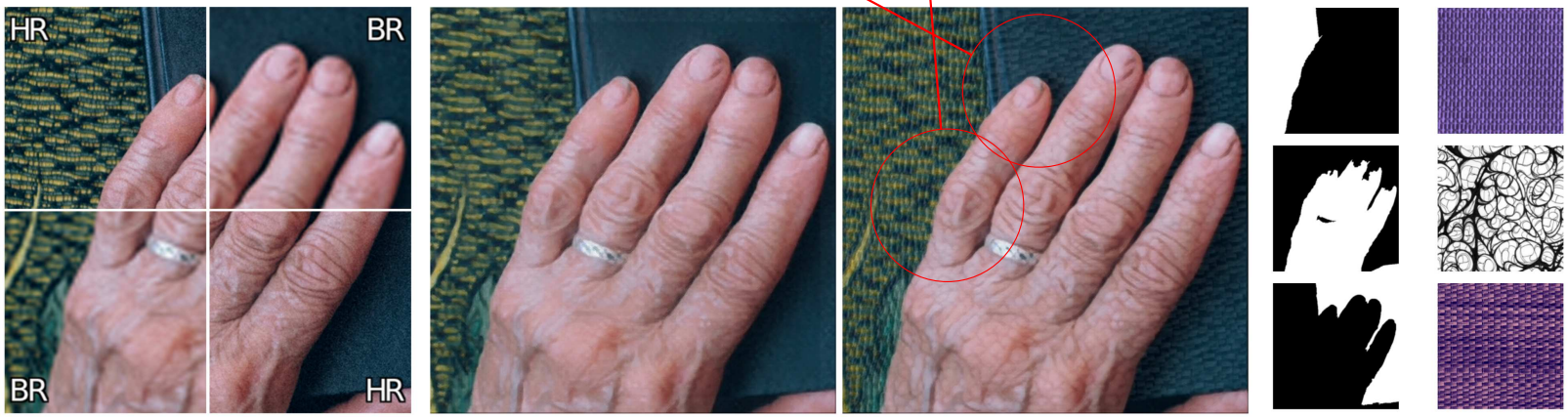
Une approche en **deux étapes** :

- Etape(s) de Stylisation (étape-ST) (un réseau par masque et par style, de **50k paramètres** ; ordre de grandeur $\sim$ **x00Ko**).
  Utilisation d'un petit réseau branché en parallèle du réseau de l'Etape de Restauration et spécialisé dans **la stylisation des détails hautes fréquences**.
  Il y'a autant d'Etapes de Stylisation que de textures différentes choisies par l'utilisateur.

Contributions - Présentation



Etapes 2



HR & BR (interp. bicubique) | Résultat de l'étape-SR (×4) | Résultat de l'étape-ST | Masques β_i | Styles Y_i

Contributions - Architecture

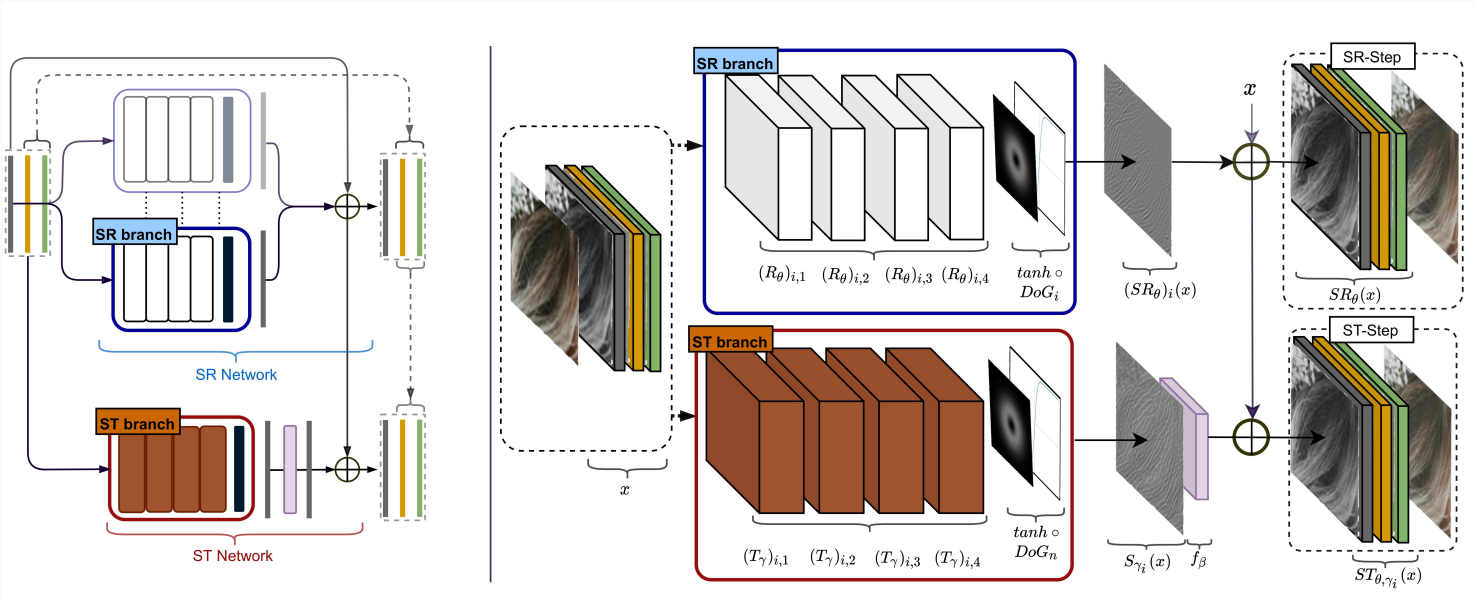


Figure 3: Architecture construite par assemblage du réseau de restauration avec le réseau de stylisation

Contributions - Architecture du réseau de restauration

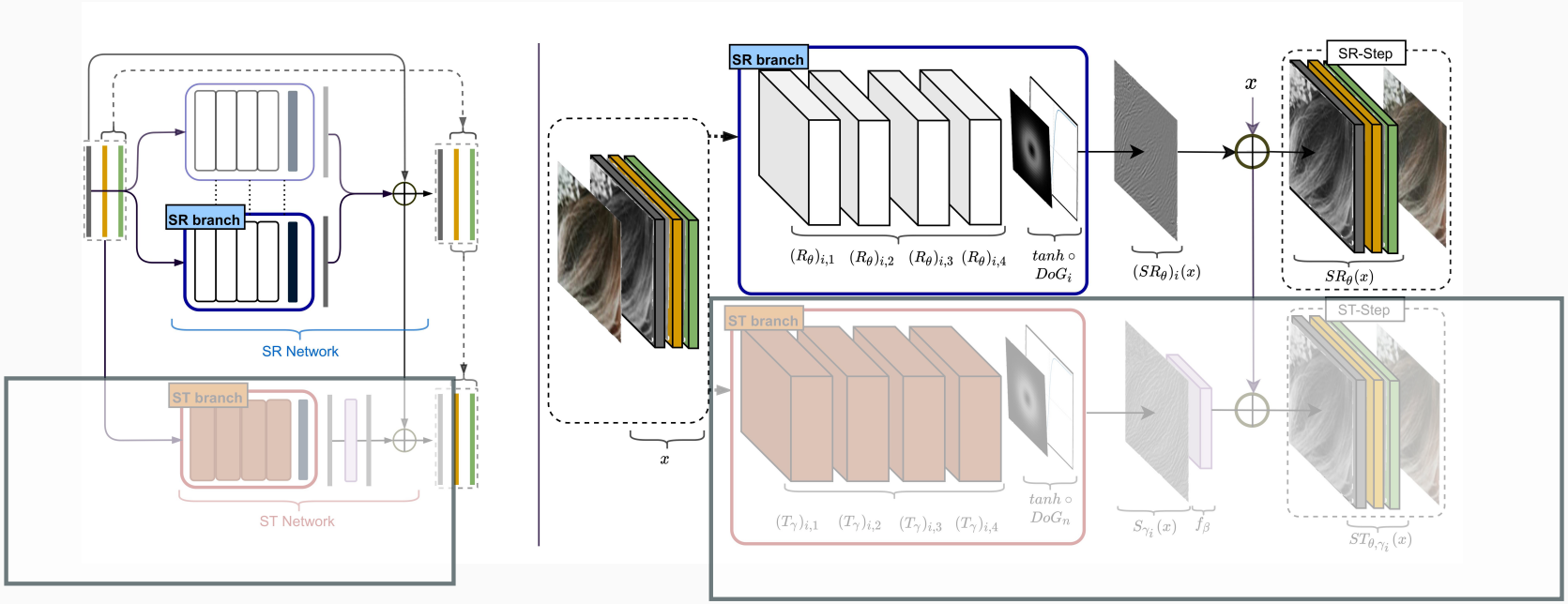


Figure 4: Architecture du Réseau de Restauration

Etape de Restauration : Architecture constituée de 6 branches parallèles. Chaque branche se termine par une différence de Gaussiennes (filtre passe-bande) spécialisant la branche dans une bande de fréquences donnée. La reconstruction des branches est une combinaison linéaire des contributions. **($\sim 200k$) paramètres**

Contributions - Entrainement du réseau de restauration

Avec θ les paramètres du réseau de restauration M_θ, il s'agit de résoudre :

$$\min_\theta \mathcal{L}_M(X, M_\theta(x)),$$

avec

$$\mathcal{L}_M(X, Y) = \sum_{k=1}^{K} \|X_k - Y_k\|^2 + \lambda_M \mathcal{L}_{Perc}(X_k, Y_k).$$

où $\|.\|$ est la norme de Frobenius,
$\mathcal{L}_{Perc}(x, y) = \sum_{\ell \in L_{Perc}} \|\phi_\ell(x) - \phi_\ell(y)\|^2$ correspond au terme de 'fidélité', avec $\phi_\ell(.)$ les caractéristiques normalisées de la couche ℓ du VGG-16[6] ($L_{Perc} = \{5, 9, 13\}$).

[6] Karen Simonyan and Andrew Zisserman (2015). "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *International Conference on Learning Representations*.

Contributions - évaluation étape de restauration

Model	# Params.	DIV2K	Set5	Set14	Bsd100
Ours	200K	1.02	2.22	1.36	0.76
SRCNN[7]	440K	0.72	1.95	0.81	0.54
EDSR[8]	1517K	1.54	4.71	1.86	1.24
SRGAN[9]	1554K	-0.50	2.00	-0.19	-0.48
RDN[10]	2205K	1.59	4.76	1.83	1.29

Figure 5: comparaison des Gains en PSNR entre notre réseau de restauration (étape-1 ; entrainé avec l'erreur quadratique moyenne comme seule pénalité, *i.e* $\lambda_M = 0$) et d'autres réseaux

[7] Chao Dong et al. (2014). "Learning a Deep Convolutional Network for Image Super-Resolution". In: *Computer Vision – ECCV 2014*. Ed. by David Fleet et al. Cham: Springer International Publishing, pp. 184–199. ISBN: 978-3-319-10593-2.

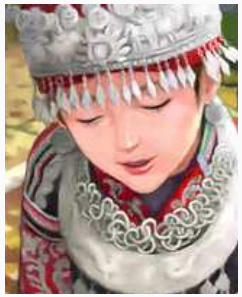
[8] Bee Lim et al. (2017). "Enhanced Deep Residual Networks for Single Image Super-Resolution". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1132–1140. DOI: 10.1109/CVPRW.2017.151.

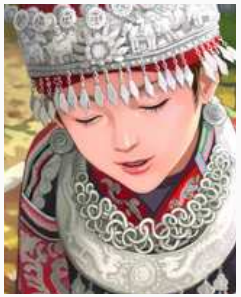
[9] Christian Ledig et al. (July 2017). "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network". In: pp. 105–114. DOI: 10.1109/CVPR.2017.19.

[10] Yulun Zhang et al. (June 2018). "Residual Dense Network for Image Super-Resolution". In: pp. 2472–2481. DOI: 10.1109/CVPR.2018.00262.

Contributions - évaluation étape de restauration


x_k (bruitée)


Etape 1


X_k originale


x_k (floutée)

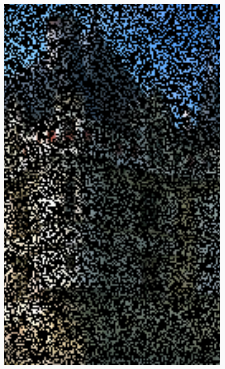

Etape 1


X_k originale


x_k (basse-rés.)


Etape 1


X_k originale


x_k (inpainting)


Etape 1


X_k originale

Contributions - Architecture étape de stylisation

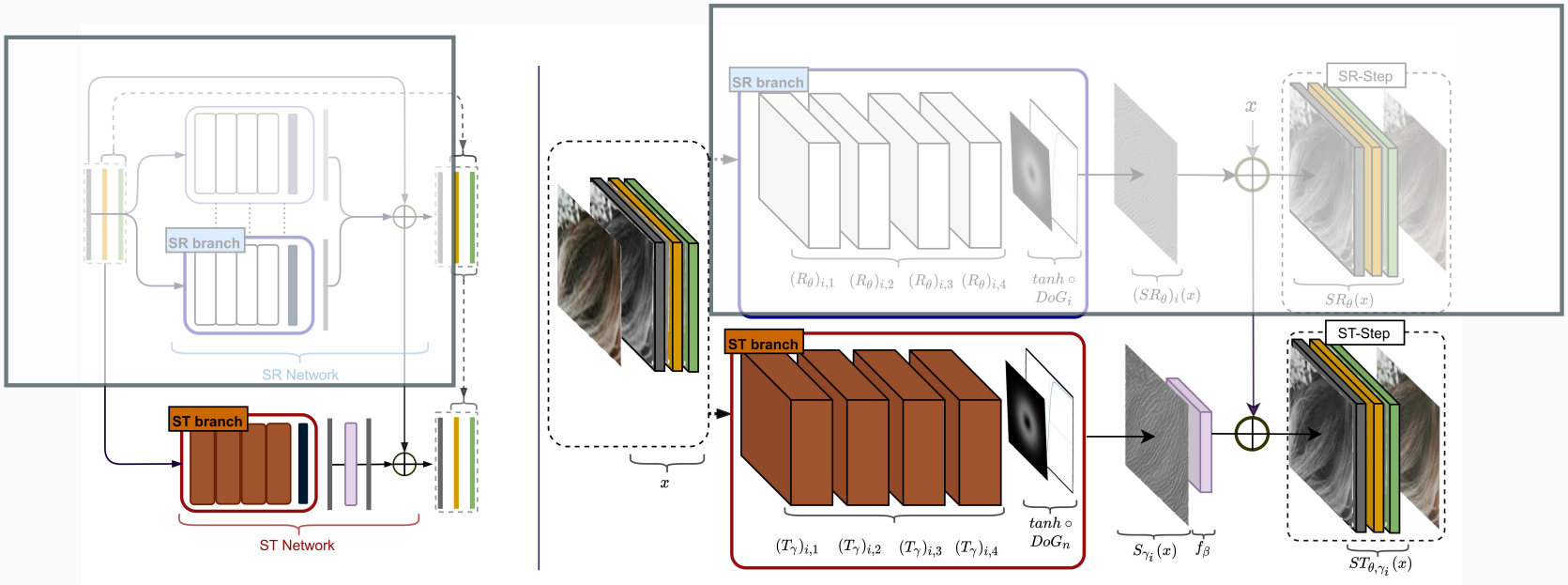


Figure 6: Architecture du Réseau de Stylisation

Etapes-ST : Architecture très légère ($\sim$ **50k paramètres**) en parallèle du réseau de restauration dont les paramètres sont gelés. Le reconstruction est additive. A la fin du réseau-ST, un filtre passe-haut suivi d'une normalisation par batch dont les paramètres affines sont contrôlés permet la génération d'un résidu haute fréquence, globalement nul et à écart type contrôlé. Un réseau-ST par style.

Contributions - Entrainement du réseau de stylisation

Avec θ les paramètres du réseau de restauration M_θ (paramètres gelés) et γ_i les paramètres du réseau i de sylisation ST_{θ,γ_i} (associé au style Y_i), il s'agit de résoudre : $\min_{\gamma_i} \mathcal{L}_{\mathsf{ST}}\left(\mathsf{X}, Y_i, \mathsf{ST}_{\theta,\gamma_i}(\mathsf{x})\right)$
avec

$$\mathcal{L}_{\mathsf{ST}}(\mathsf{X}, Y_i, \mathsf{Z}) = \sum_{k=1}^{K} \lambda_{ST} \mathcal{L}_{\mathsf{Perc}}(X_k, Z_k) + \mathcal{L}_{\mathsf{Tex}}(Y_i, Z_k).$$

où la pénalité de texture est définie à partir des matrices de gram normalisées G,[11]

$$\mathcal{L}_{\mathsf{Tex}}(x, y) = \sum_{\ell \in L_{\mathsf{Tex}}} \|G(\phi_\ell(x)) - G(\phi_\ell(y))\|^2. \tag{1}$$

$(\mathsf{L}_{\mathsf{Tex}} = \{2, 5, 9\}$, $\mathsf{L}_{\mathsf{Perc}} = \{7\})$

[11] Leon Gatys, Alexander Ecker, and Matthias Bethge (Aug. 2015). "A Neural Algorithm of Artistic Style". In: *arXiv*. DOI: 10.1167/16.12.326.

Contributions - résultats étapes de stylisation

Image originale X_k | Image dégradée x_k | Etape 1 ('restauration') | Etape 2 (stylisation)

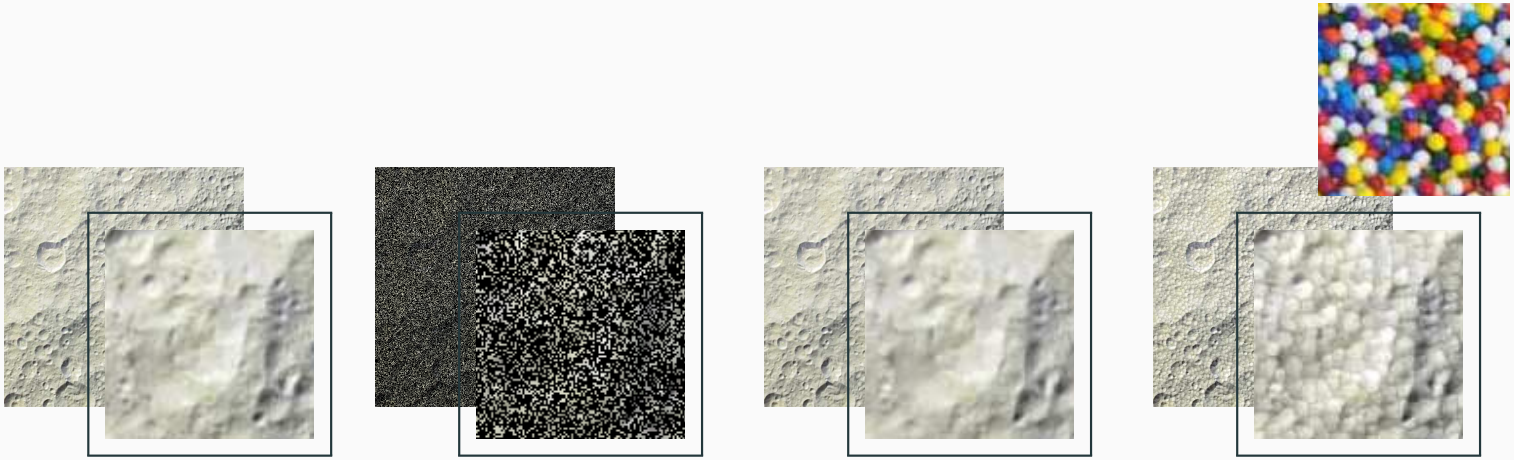


Image originale X_k | Image dégradée x_k | Etape 1 ('restauration') | Etape 2 (stylisation)

Contributions - résultats étapes de stylisation

Image originale X_k Image dégradée x_k Etape 1 ('restauration') Etape 2 (stylisation)

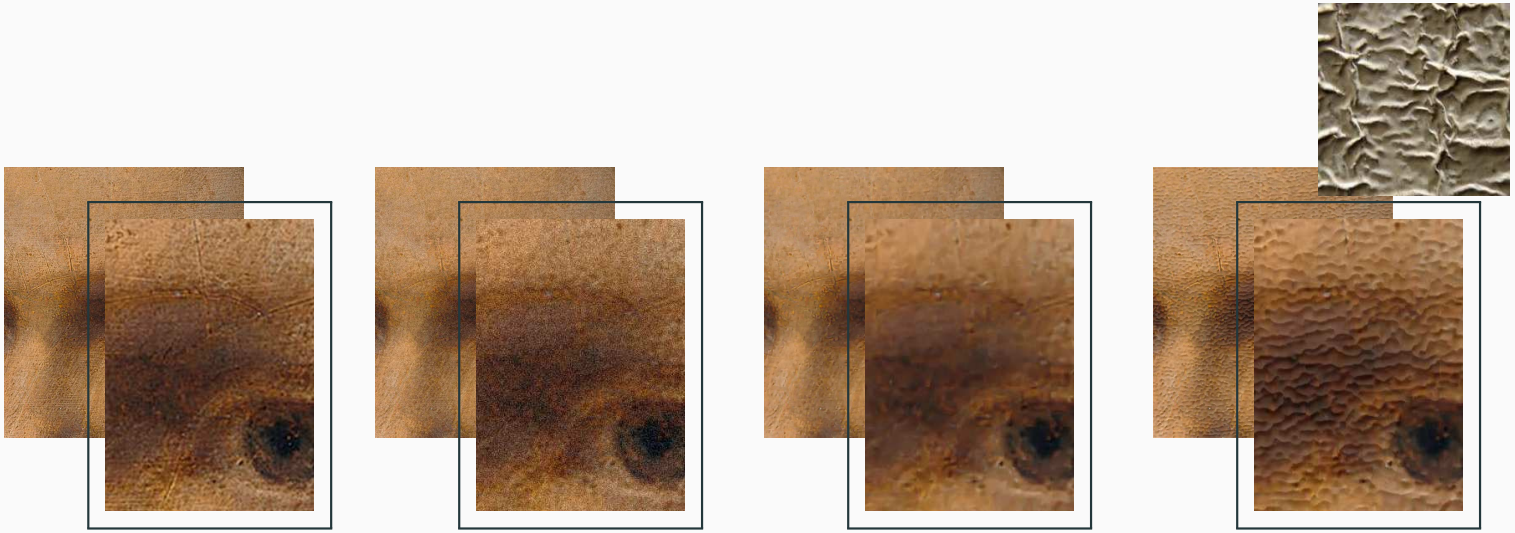


Image originale X_k Image dégradée x_k Etape 1 ('restauration') Etape 2 (stylisation)

Contributions - résultats étape de stylisation

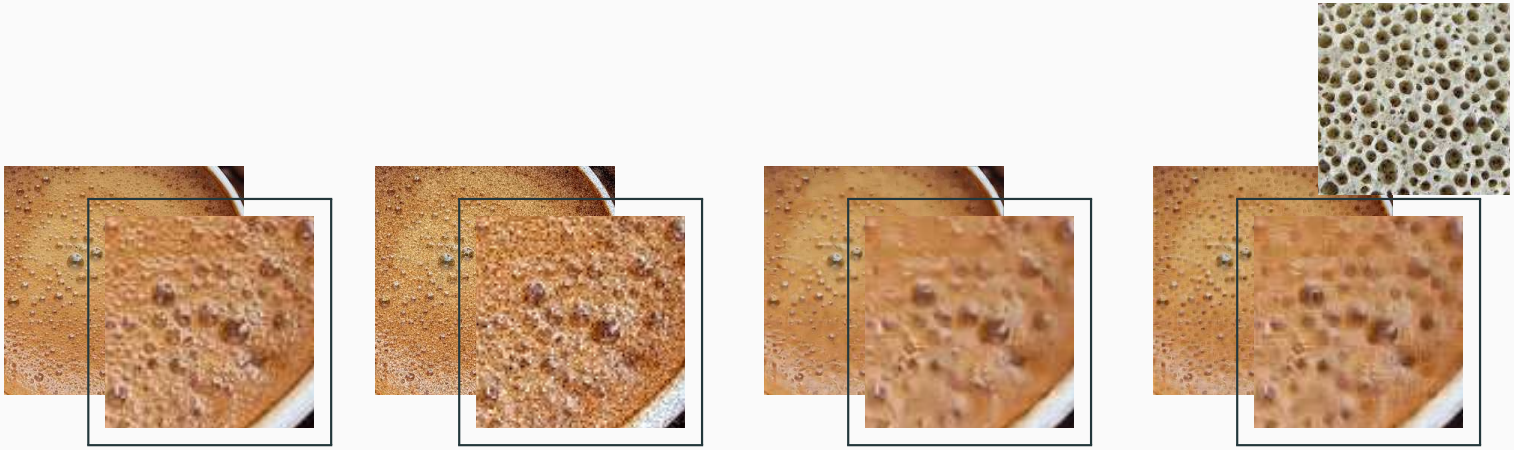


Image originale X_k | Image dégradée x_k | Etape 1 ('restauration') | Etape 2 (stylisation)

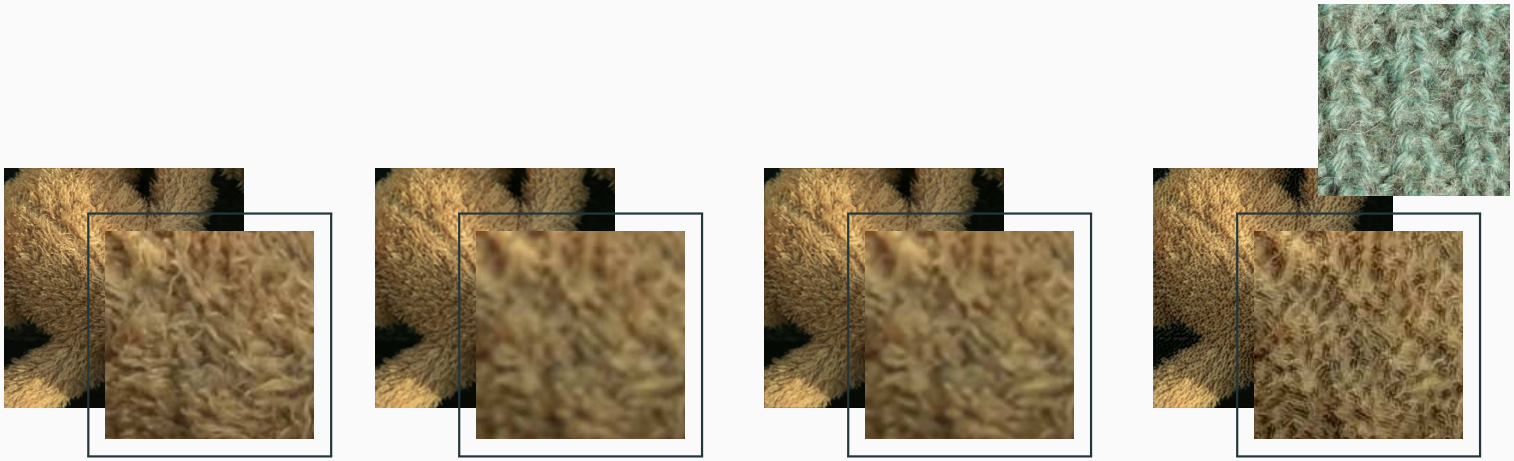
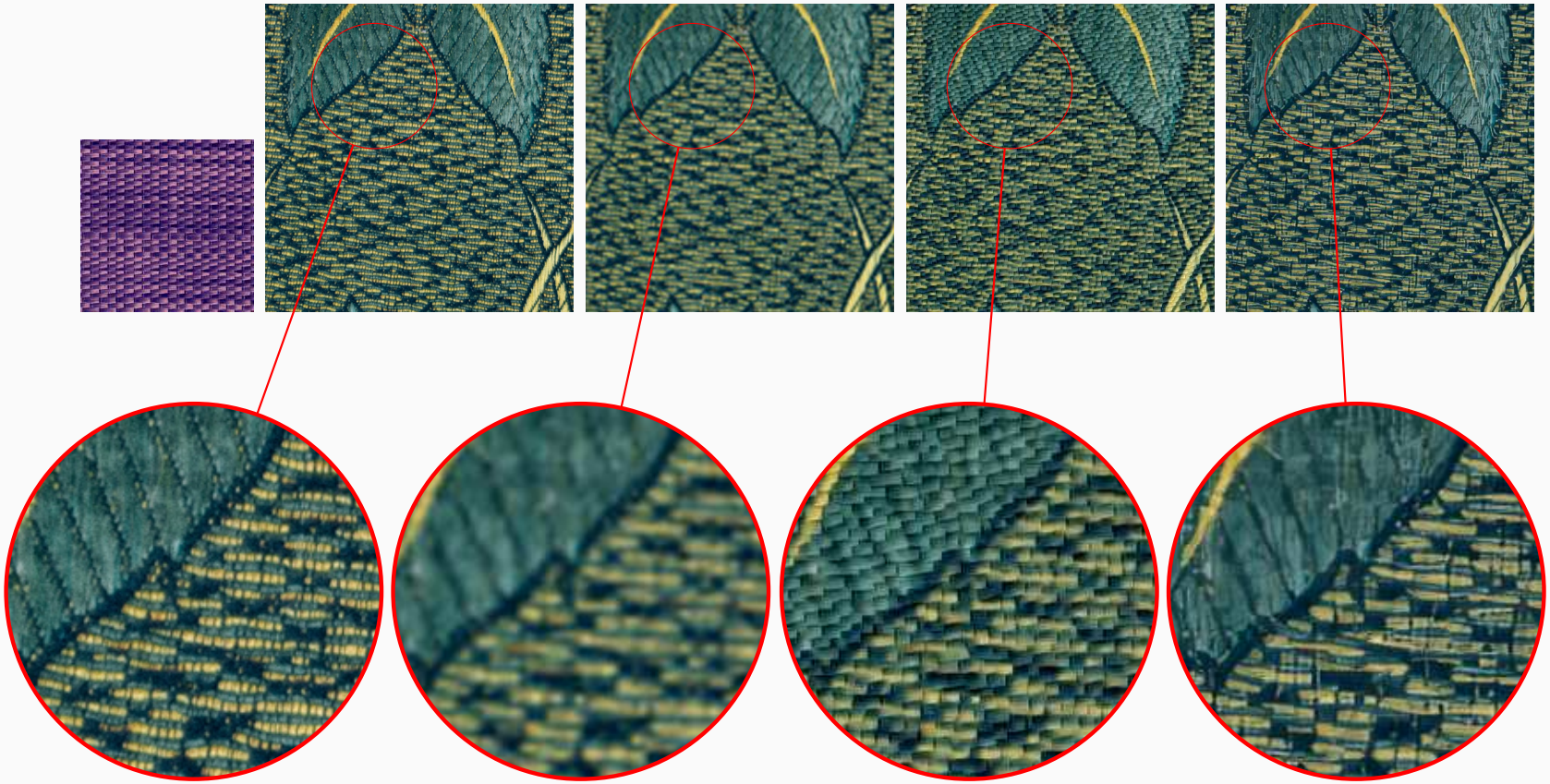


Image originale X_k | Image dégradée x_k | Etape 1 ('restauration') | Etape 2 (stylisation)

Contributions - comparaison état de l'art (Reference-based SR)

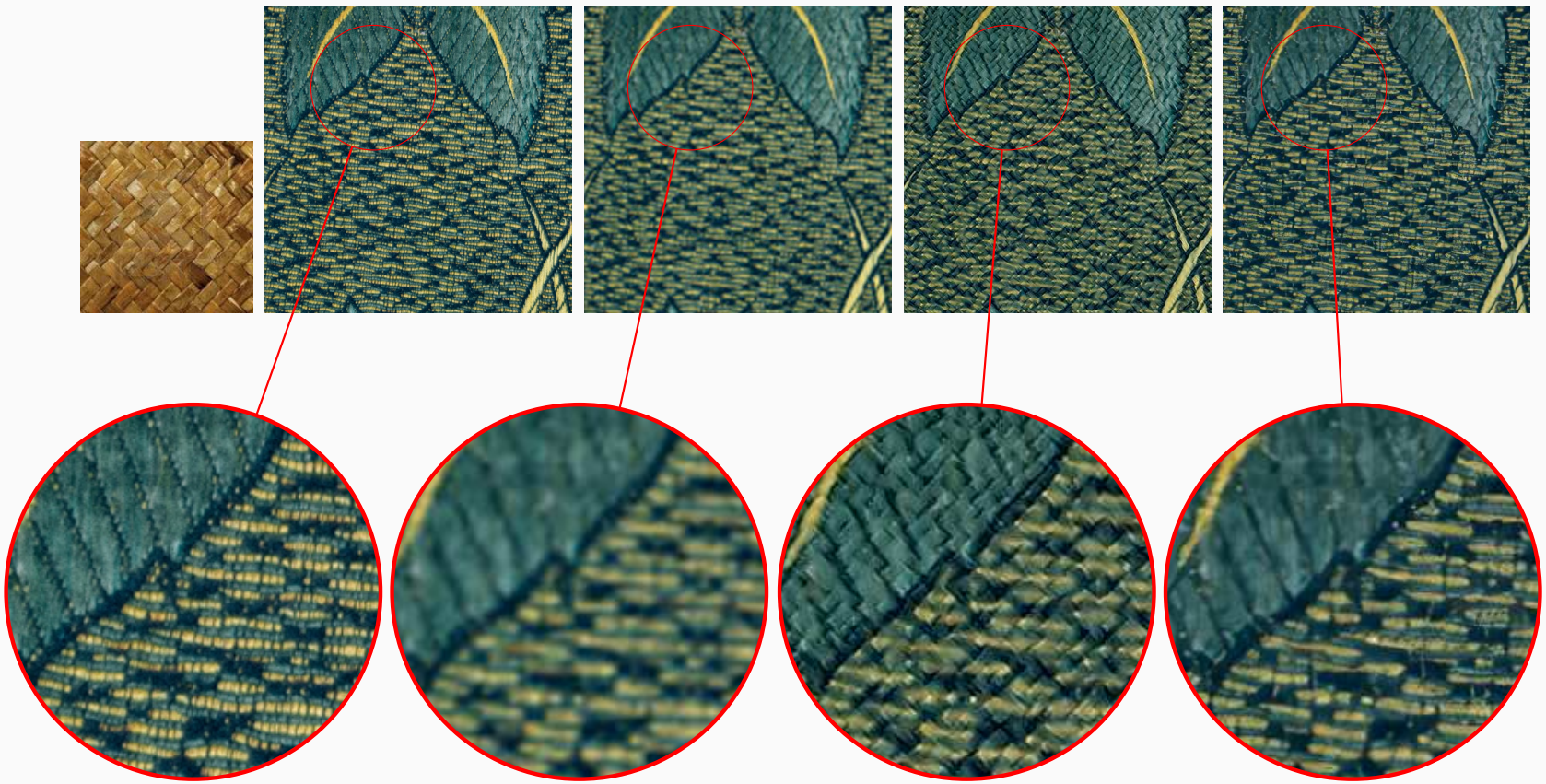
Style ; Haute-Rés. ; Basse-Rés. & SR stylisée (250k params) & Résultat TTSR[12] (9000k params)



[12]Fuzhi Yang et al. (2020). "Learning Texture Transformer Network for Image Super-Resolution". In: *CVPR*.

Contributions - comparaison état de l'art (Reference-based SR)

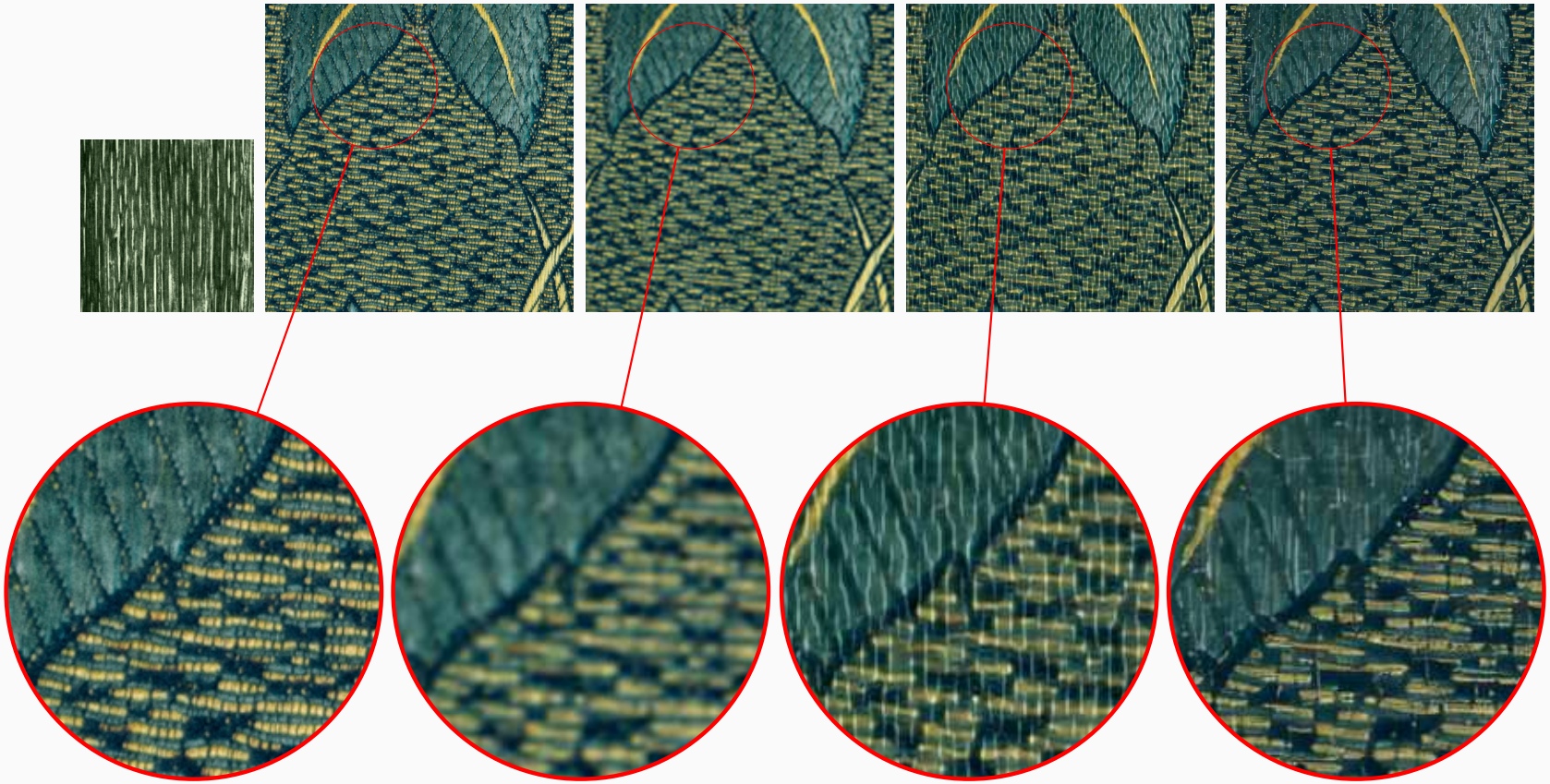
Style ; Haute-Rés. ; Basse-Rés. & SR stylisée (250k params) & Résultat TTSR[13] (9000k params)



[13]Fuzhi Yang et al. (2020). "Learning Texture Transformer Network for Image Super-Resolution". In: *CVPR*.

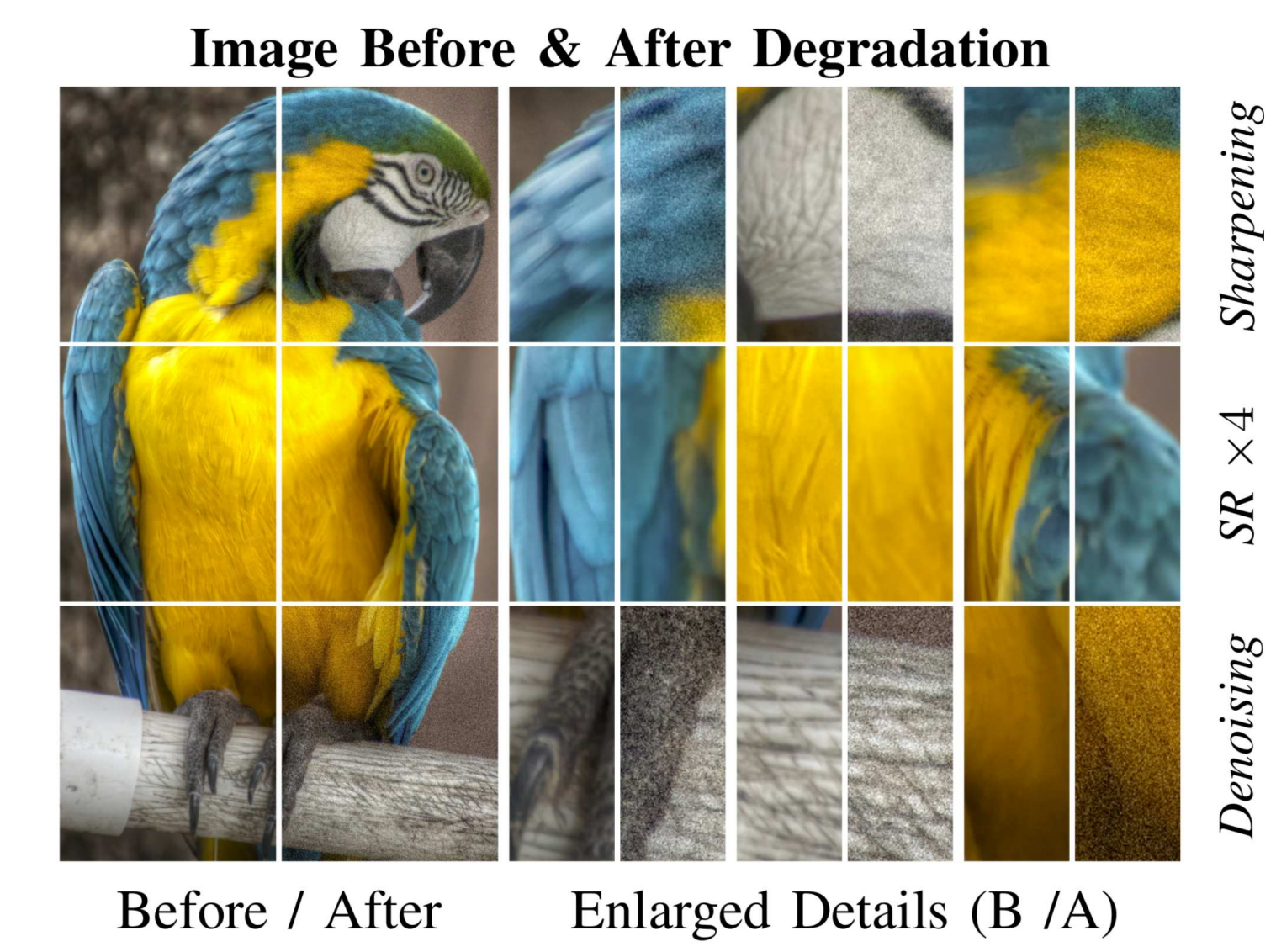
Contributions - comparaison état de l'art (Reference-based SR)

Style ; Haute-Rés. ; Basse-Rés. & SR stylisée (250k params) & Résultat TTSR[14] (9000k params)

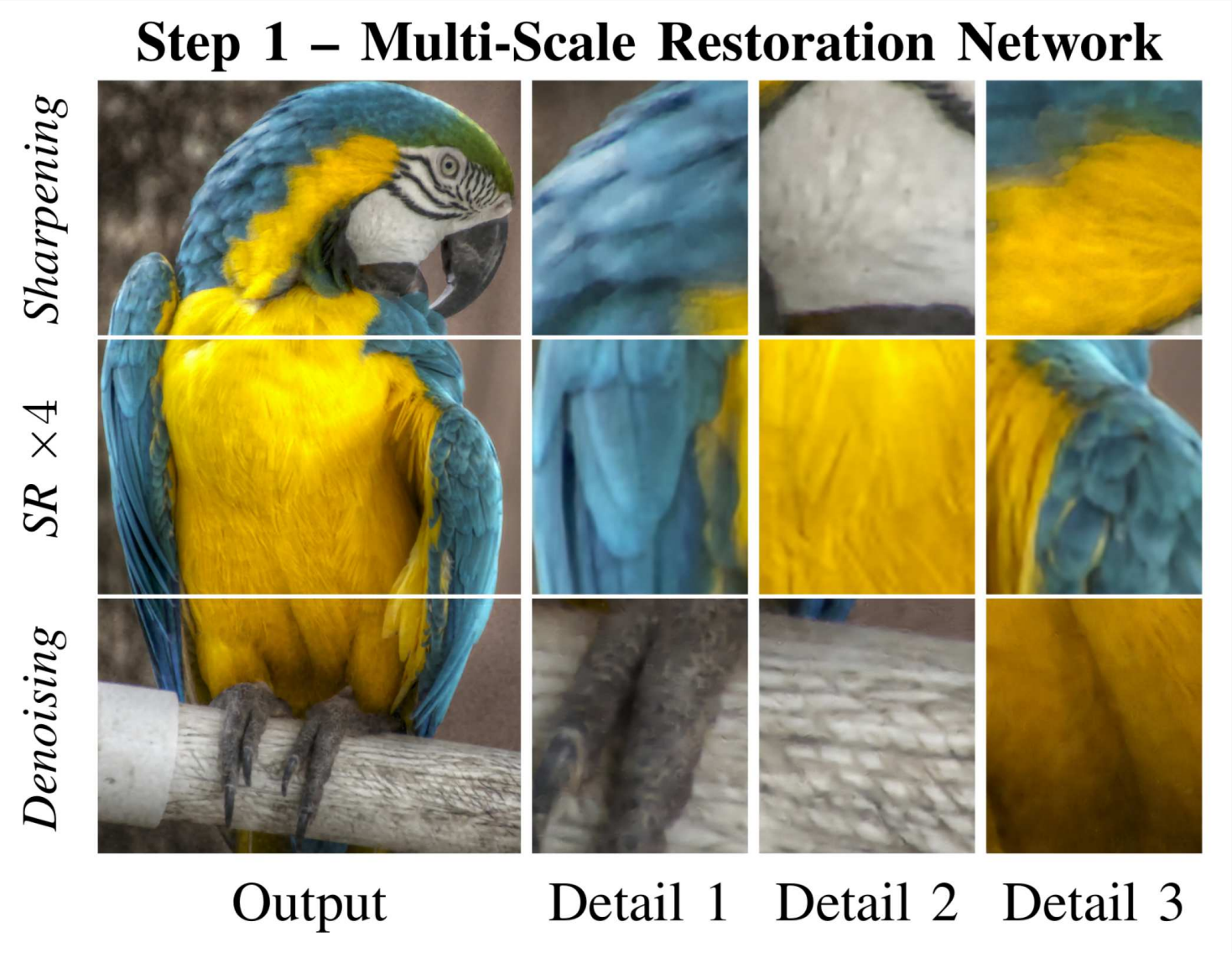


[14]Fuzhi Yang et al. (2020). "Learning Texture Transformer Network for Image Super-Resolution". In: *CVPR*.

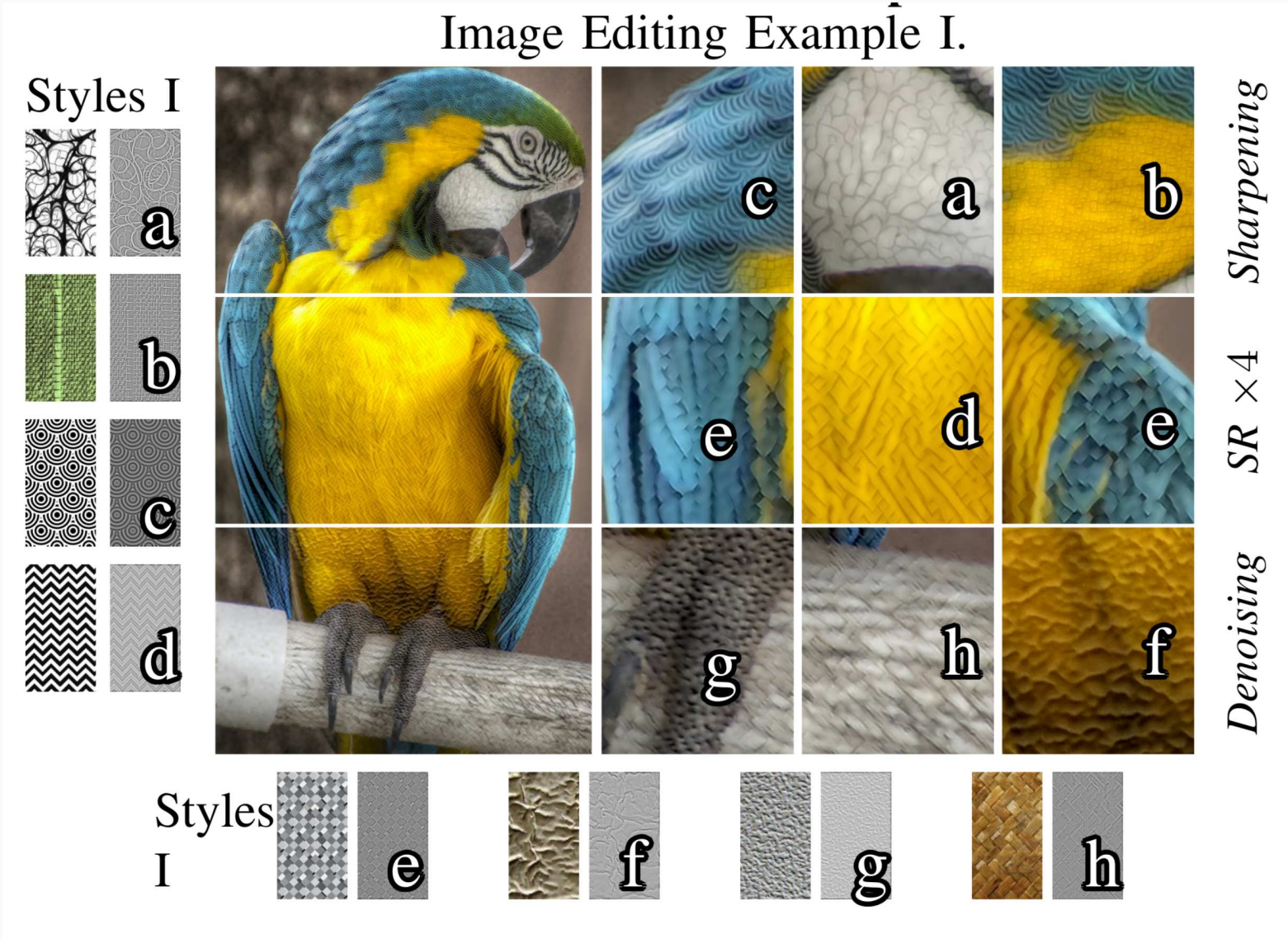
Exemple image multi-textures



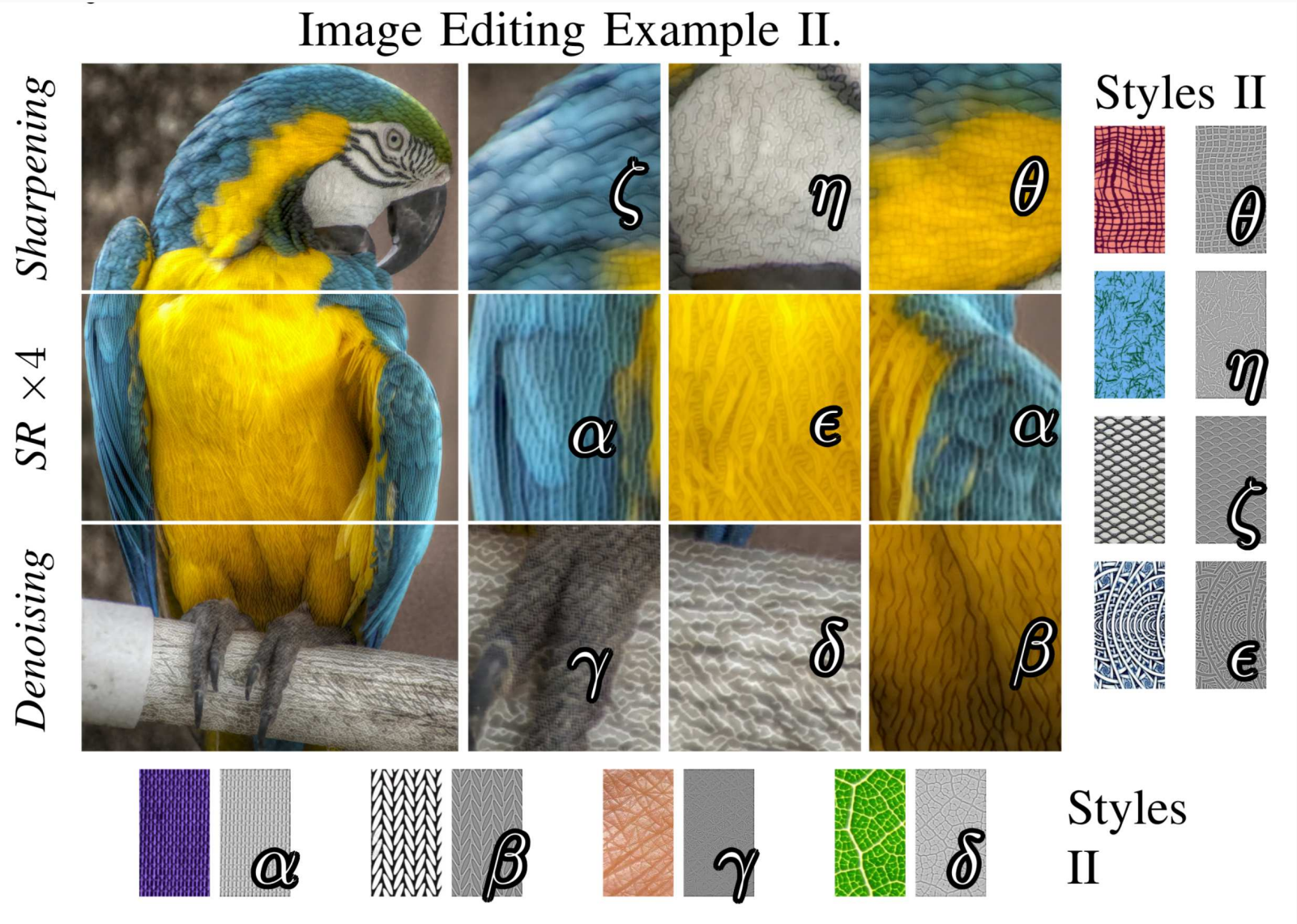
Exemple image multi-textures



Exemple image multi-textures



Exemple image multi-textures



Merci pour Votre attention

Merci pour votre attention, des questions ?

References i

Dong, Chao et al. (2014). “Learning a Deep Convolutional Network for Image Super-Resolution”. In: *Computer Vision – ECCV 2014*. Ed. by David Fleet et al. Cham: Springer International Publishing, pp. 184–199. ISBN: 978-3-319-10593-2.

Gatys, Leon, Alexander Ecker, and Matthias Bethge (Aug. 2015). “A Neural Algorithm of Artistic Style”. In: *arXiv*. DOI: 10.1167/16.12.326.

References ii


Johnson, Justin, Alexandre Alahi, and Li Fei-Fei (Oct. 2016). "Perceptual Losses for Real-Time Style Transfer and Super-Resolution". In: vol. 9906, pp. 694–711. ISBN: 978-3-319-46474-9. DOI: 10.1007/978-3-319-46475-6_43.


Ledig, Christian et al. (July 2017). "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network". In: pp. 105–114. DOI: 10.1109/CVPR.2017.19.


Lim, Bee et al. (2017). "Enhanced Deep Residual Networks for Single Image Super-Resolution". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1132–1140. DOI: 10.1109/CVPRW.2017.151.

References iii

Simonyan, Karen and Andrew Zisserman (2015). "Very Deep Convolutional Networks for Large-Scale Image Recognition". In: *International Conference on Learning Representations*.

Wang, Xintao et al. (Sept. 2018). *ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks*.

Yang, Fuzhi et al. (2020). "Learning Texture Transformer Network for Image Super-Resolution". In: *CVPR*.

Zhang, Yulun et al. (June 2018). "Residual Dense Network for Image Super-Resolution". In: pp. 2472–2481. DOI: 10.1109/CVPR.2018.00262.

Bi-Scale Style Transfer

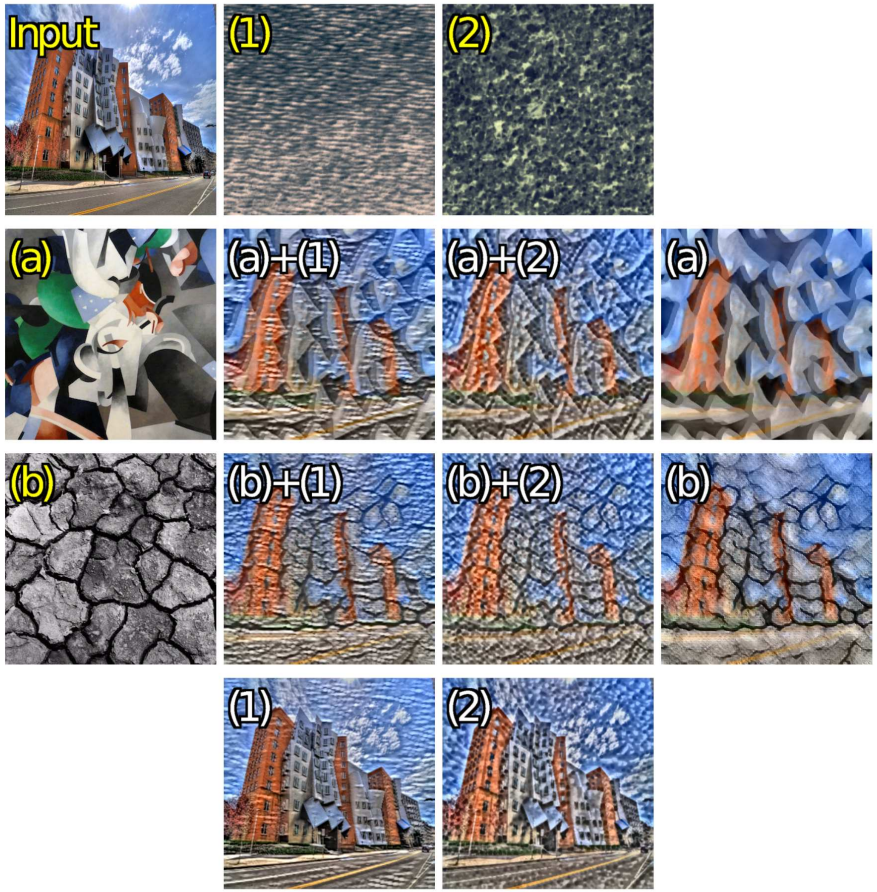


FIGURE 1 – *Résultats du transfert de styles à deux échelles. L'image de contenu* Input *est stylisée avec les styles* a *and* b *combinés aux textures* 1 *and* 2 *et ce à l'aide de réseaux légers et modulables. Le transfert de style pour les réseaux pris individuellement correspond à* a, b, 1, 2.