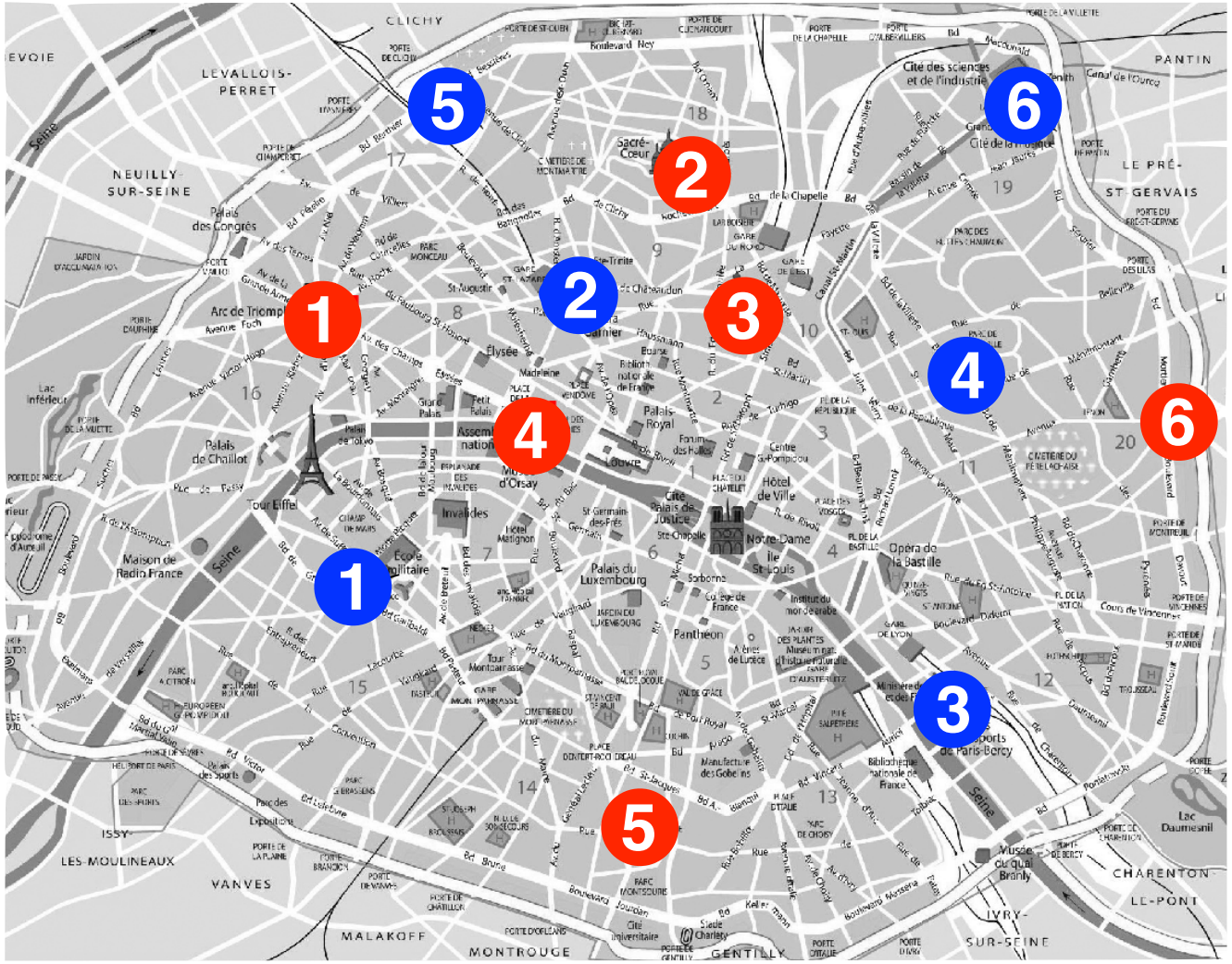
LA STRATÉGIE DU MOINDRE EFFORT POUR APPRENDRE AUX MACHINES

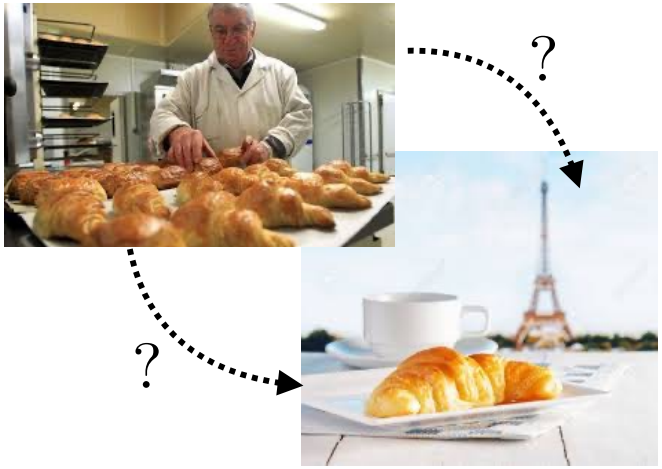
Transport optimal et intelligence artificielle

Gabriel Peyré & Rémi Flamary

30 Mars 2023, Jussieu

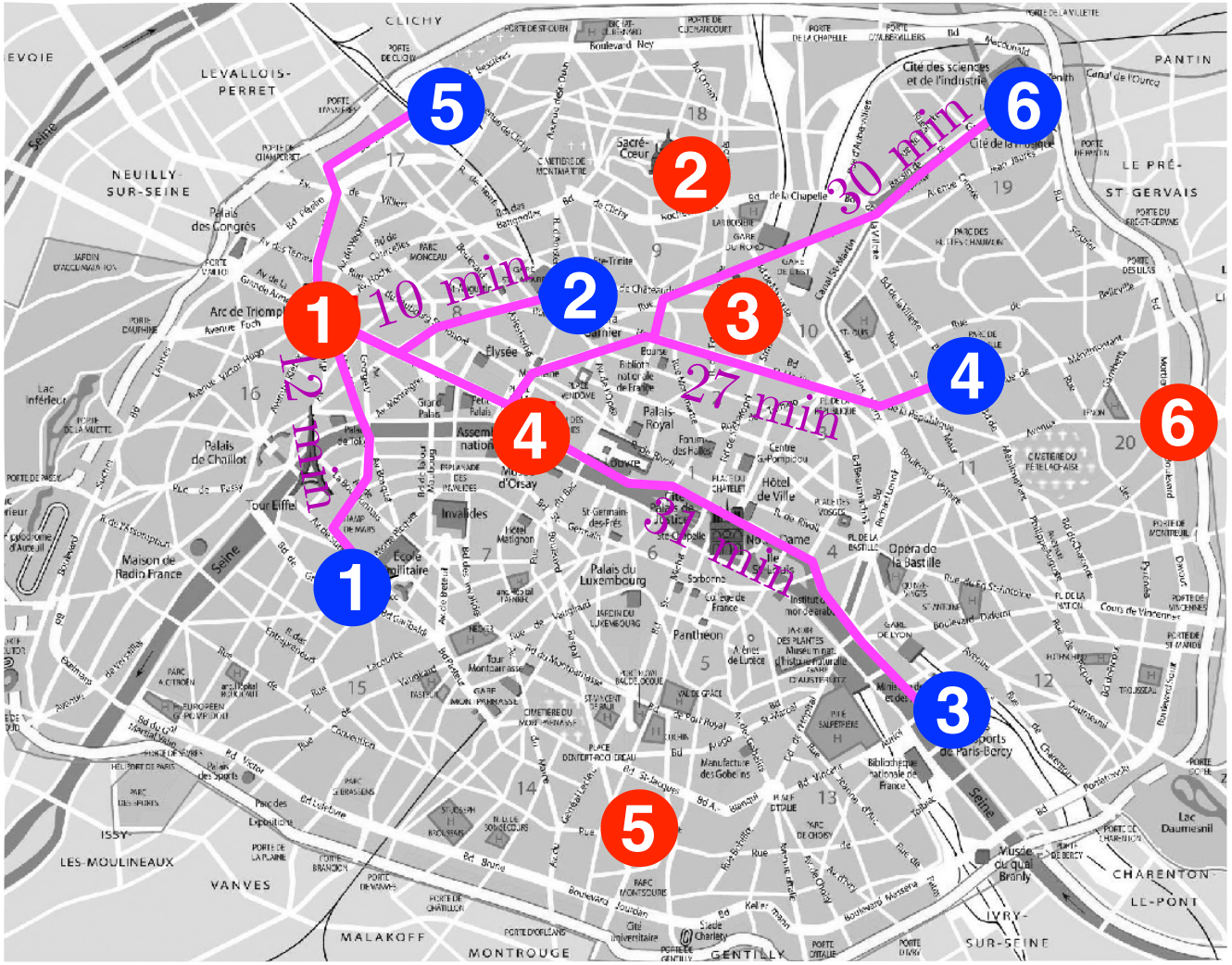
Des boulangeries aux cafés





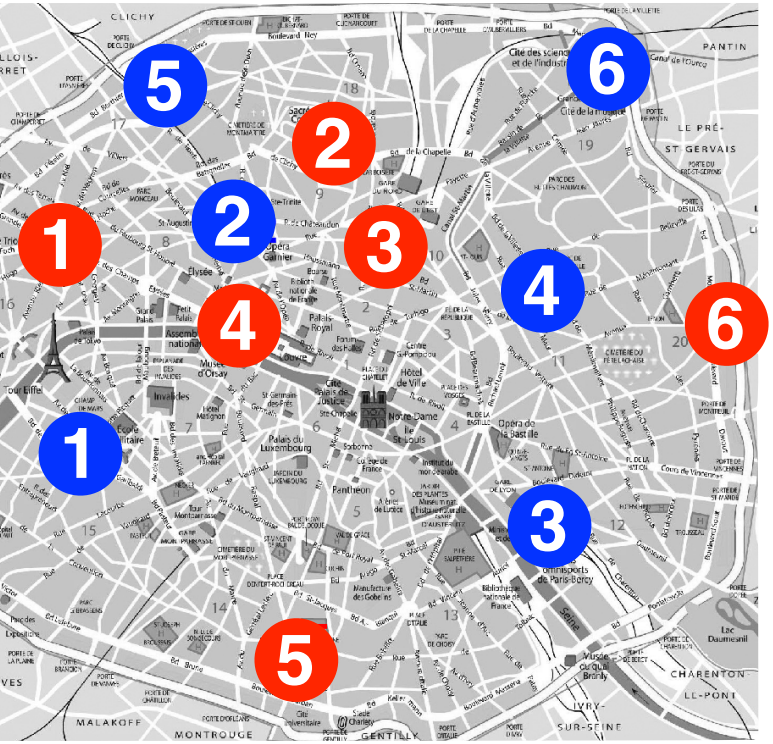
	1	2	3	4	5	6
1	12	10	31	27	10	30
2	22	7	25	15	11	14
3	19	7	19	10	15	15
4	10	6	21	19	14	24
5	15	23	14	24	31	34
6	35	26	16	9	34	15

Des boulangeries aux cafés



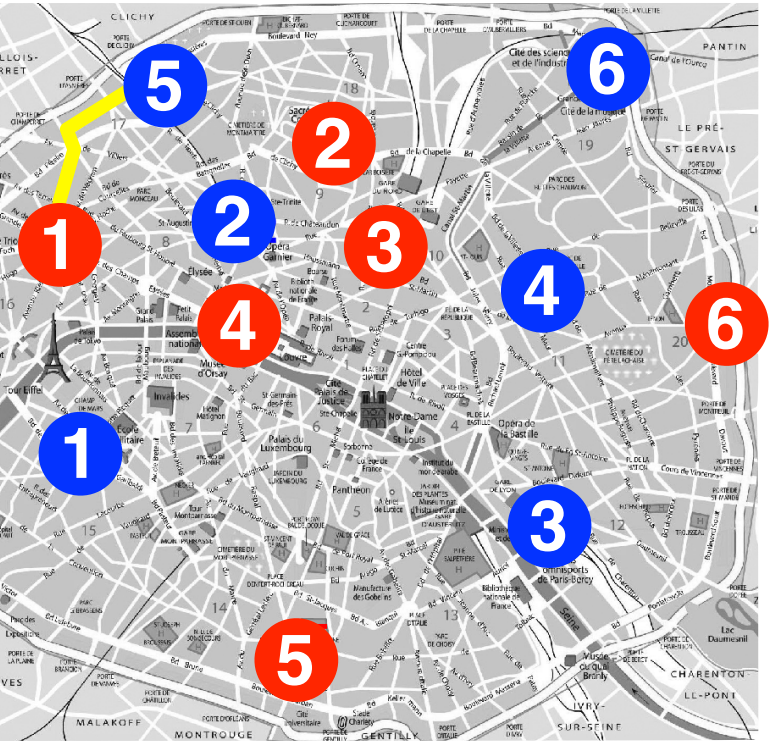
	1	2	3	4	5	6
1	12	10	31	27	10	30
2	22	7	25	15	11	14
3	19	7	19	10	15	15
4	10	6	21	19	14	24
5	15	23	14	24	31	34
6	35	26	16	9	34	15

Ma proposition



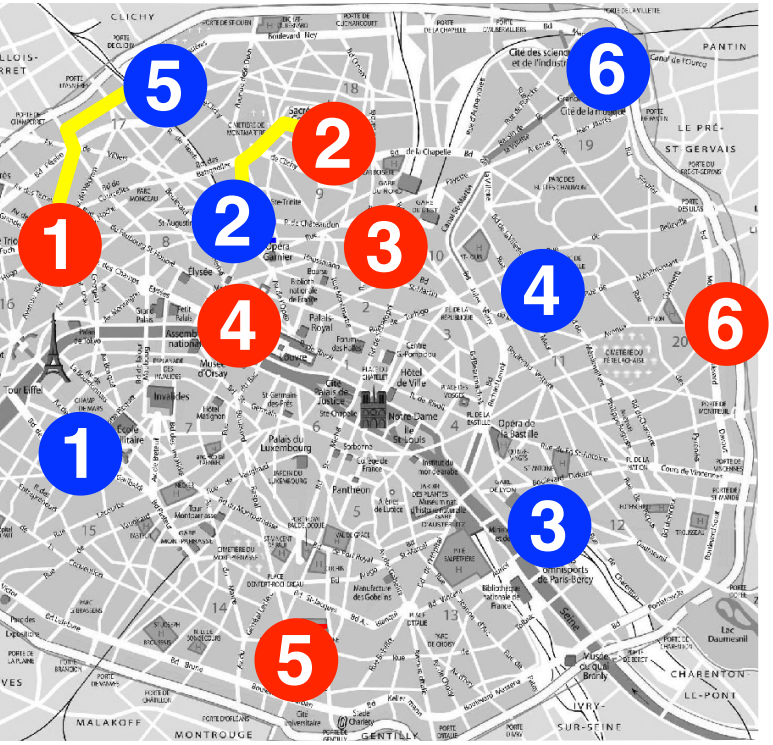
	1	2	3	4	5	6
1	12	10	31	27	10	30
2	22	7	25	15	11	14
3	19	7	19	10	15	15
4	10	6	21	19	14	24
5	15	23	14	24	31	34
6	35	26	16	9	34	15

Ma proposition



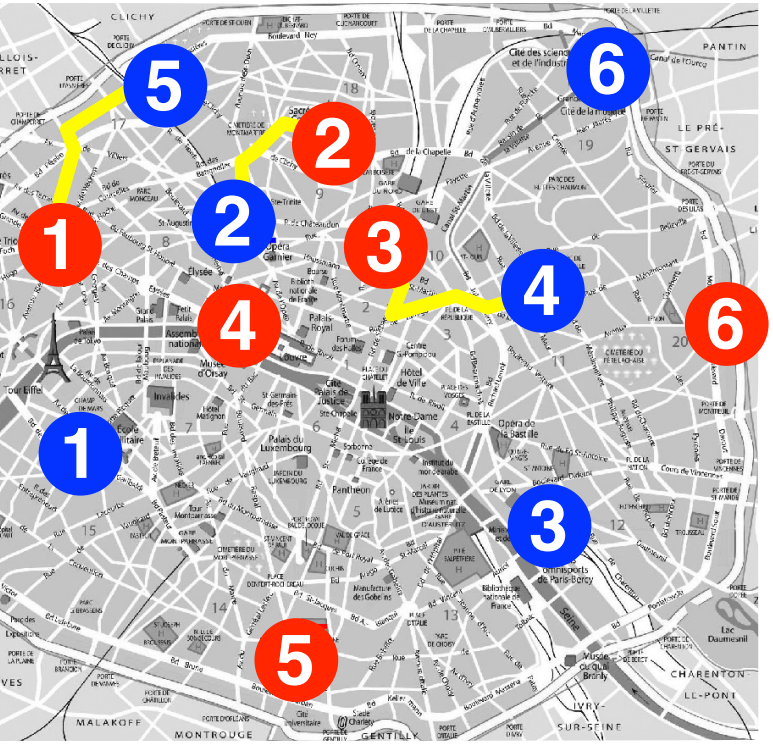
	1	2	3	4	5	6
1	12	10	31	27	10	30
2	22	7	25	15	11	14
3	19	7	19	10	15	15
4	10	6	21	19	14	24
5	15	23	14	24	31	34
6	35	26	16	9	34	15

Ma proposition



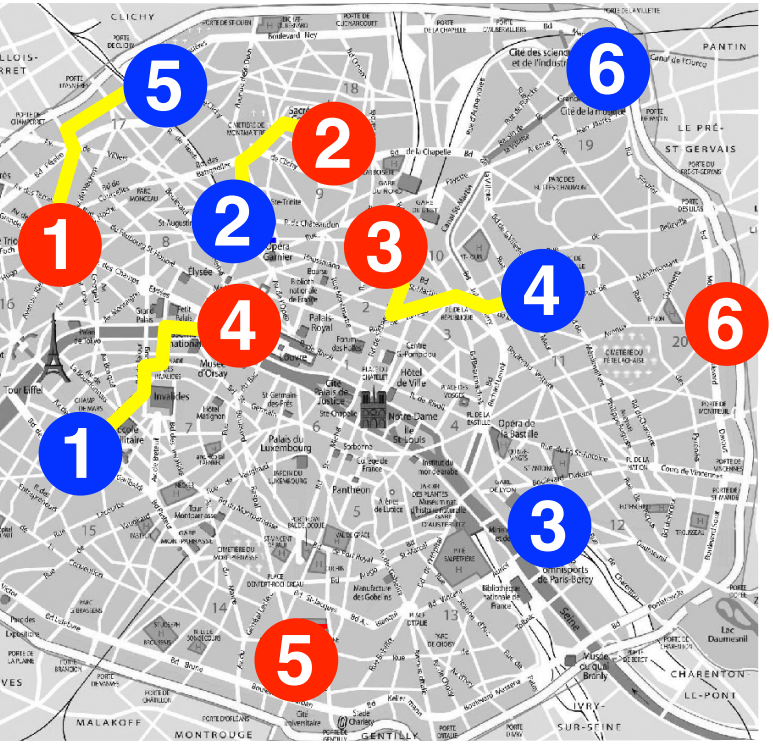
	1	2	3	4	5	6
1	12	10	31	27	10	30
2	22	7	25	15	11	14
3	19	7	19	10	15	15
4	10	6	21	19	14	24
5	15	23	14	24	31	34
6	35	26	16	9	34	15

Ma proposition



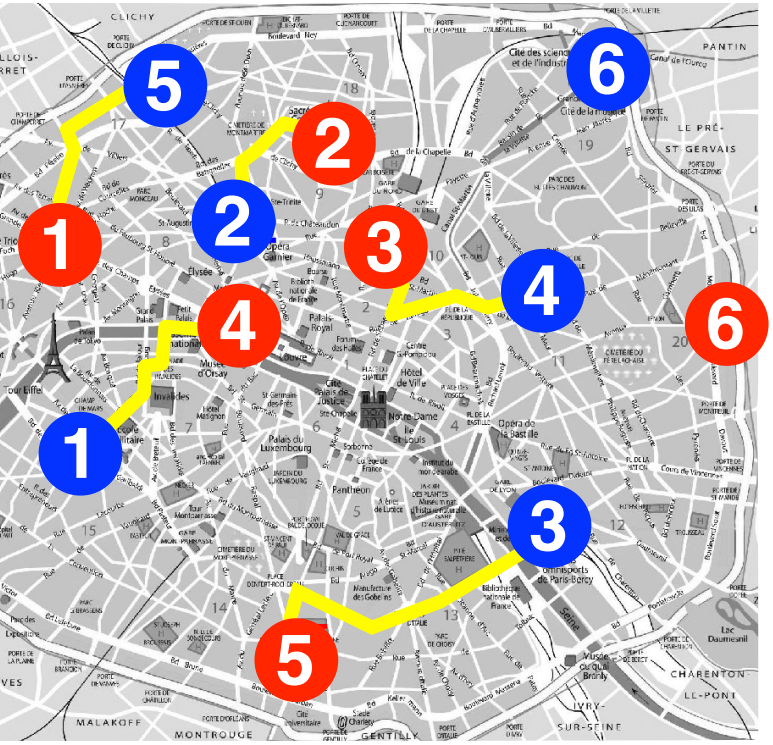
	1	2	3	4	5	6
1	12	10	31	27	10	30
2	22	7	25	15	11	14
3	19	7	19	10	15	15
4	10	6	21	19	14	24
5	15	23	14	24	31	34
6	35	26	16	9	34	15

Ma proposition



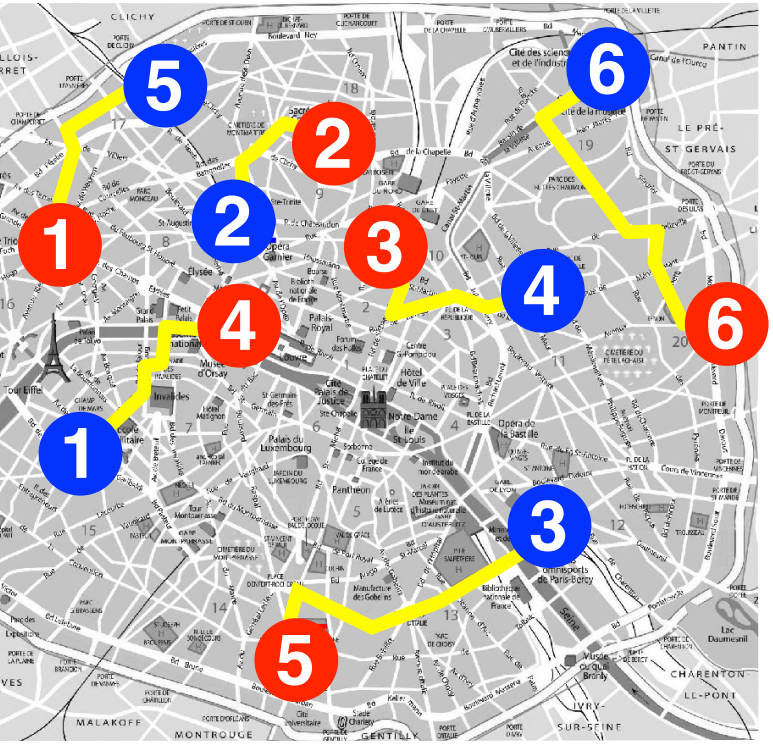
	1	2	3	4	5	6
1	12	10	31	27	10	30
2	22	7	25	15	11	14
3	19	7	19	10	15	15
4	10	6	21	19	14	24
5	15	23	14	24	31	34
6	35	26	16	9	34	15

Ma proposition



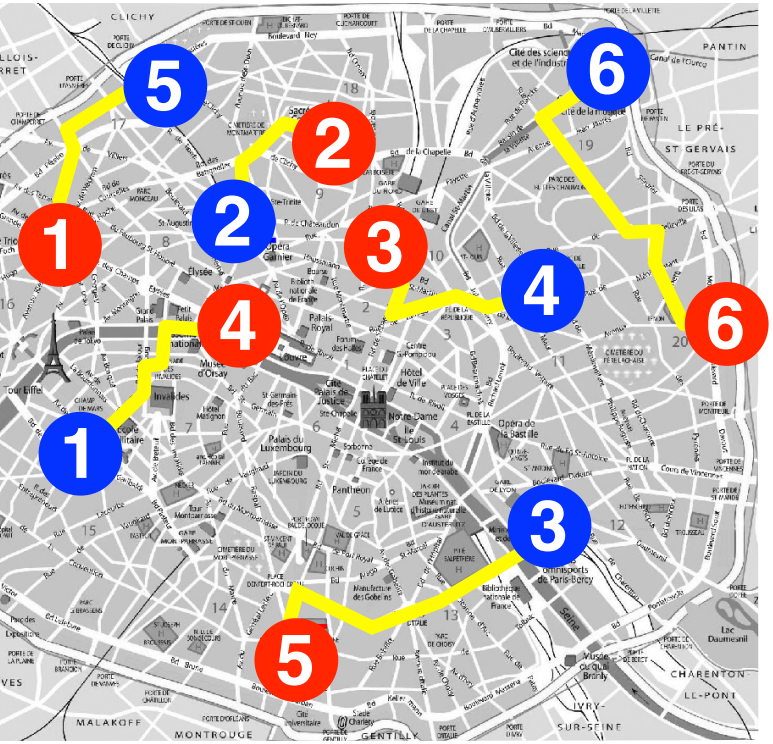
	1	2	3	4	5	6
1	12	10	31	27	10	30
2	22	7	25	15	11	14
3	19	7	19	10	15	15
4	10	6	21	19	14	24
5	15	23	14	24	31	34
6	35	26	16	9	34	15

Ma proposition



	1	2	3	4	5	6
1	12	10	31	27	10	30
2	22	7	25	15	11	14
3	19	7	19	10	15	15
4	10	6	21	19	14	24
5	15	23	14	24	31	34
6	35	26	16	9	34	15

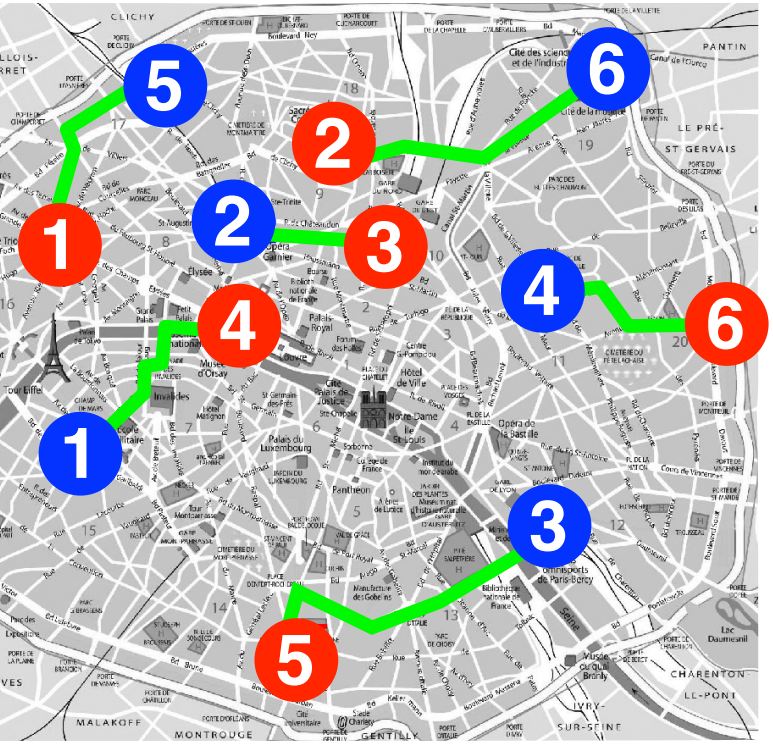
Ma proposition



	1	2	3	4	5	6
1	12	10	31	27	10	30
2	22	7	25	15	11	14
3	19	7	19	10	15	15
4	10	6	21	19	14	24
5	15	23	14	24	31	34
6	35	26	16	9	34	15

Cout = 10 + 7 + 10 + 10 + 14 +15 = 66 minutes

La meilleure solution



	1	2	3	4	5	6
1	12	10	31	27	10	30
2	22	7	25	15	11	14
3	19	7	19	10	15	15
4	10	6	21	19	14	24
5	15	23	14	24	31	34
6	35	26	16	9	34	15

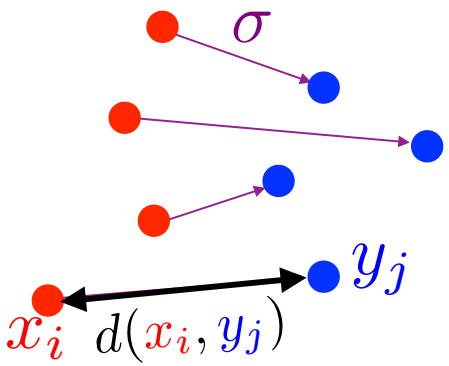
Cout = 10 + 14 + 7 + 10 + 14 + 9 = 64 minutes

Le problème de Monge : définition générale

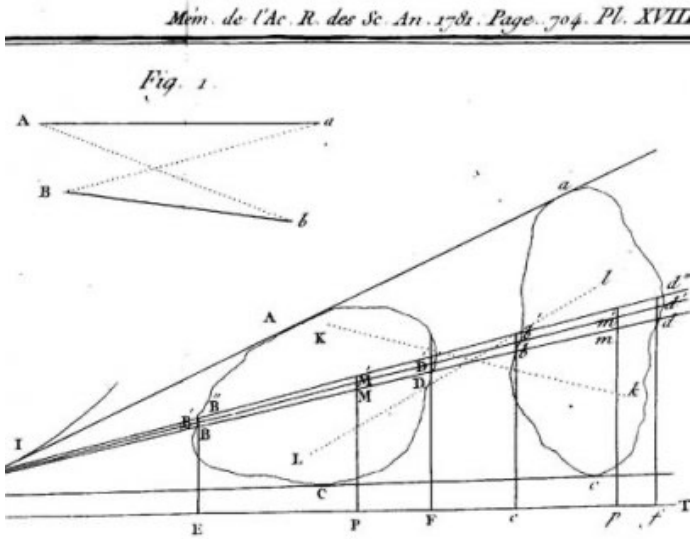
Points $(x_i)_i$, $(y_j)_j$

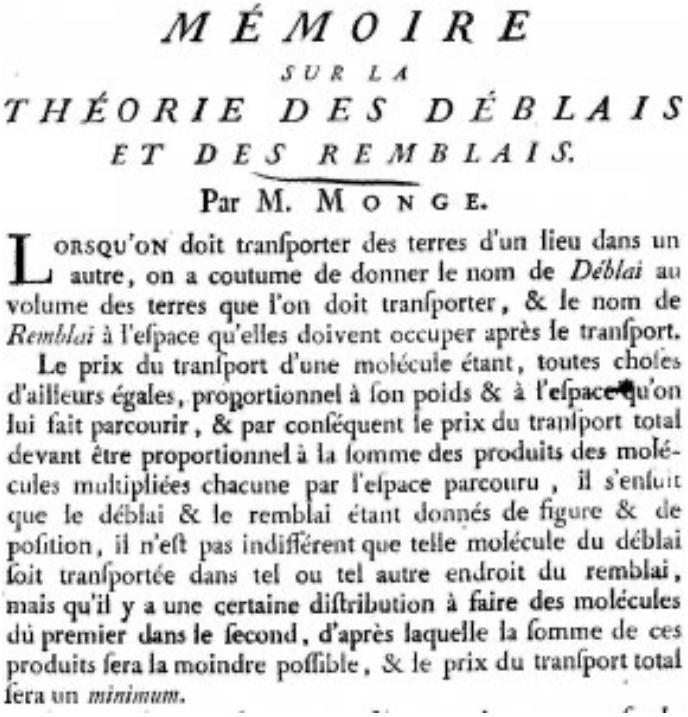
Permutation:

$$\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$$



$$\min_{\sigma} \sum_{i=1}^{n} d(x_i, y_{\sigma(i)})^p$$





MÉMOIRE
SUR LA
THÉORIE DES DÉBLAIS
ET DES REMBLAIS.
Par M. MONGE.

LORSQU'ON doit tranſporter des terres d'un lieu dans un autre, on a coutume de donner le nom de *Déblai* au volume des terres que l'on doit tranſporter, & le nom de *Remblai* à l'eſpace qu'elles doivent occuper après le tranſport.

Le prix du tranſport d'une molécule étant, toutes choſes d'ailleurs égales, proportionnel à ſon poids & à l'eſpace qu'on lui fait parcourir, & par conſéquent le prix du tranſport total devant être proportionnel à la ſomme des produits des molécules multipliées chacune par l'eſpace parcouru, il s'enſuit que le déblai & le remblai étant donnés de figure & de poſition, il n'eſt pas indifférent que telle molécule du déblai ſoit tranſportée dans tel ou tel autre endroit du remblai, mais qu'il y a une certaine diſtribution à faire des molécules du premier dans le ſecond, d'après laquelle la ſomme de ces produits ſera la moindre poſſible, & le prix du tranſport total ſera un *minimum*.

[Monge 1784]

Nombre de Possibilités

1 vers $\sigma_1 \in \{1, 2, 3, 4, 5, 6\}$.

2 vers $\sigma_2 \in \{1, \ldots, 6\} - \{\sigma_1\}$.

3 vers $\sigma_3 \in \{1, \ldots, 6\} - \{\sigma_1, \sigma_2\}$.

4 vers $\sigma_4 \in \{1, \ldots, 6\} - \{\sigma_1, \ldots, \sigma_3\}$.

5 vers $\sigma_5 \in \{1, \ldots, 6\} - \{\sigma_1, \ldots, \sigma_4\}$.

6 vers $\sigma_6 \in \{1, \ldots, 6\} - \{\sigma_1, \ldots, \sigma_5\}$.

Total : $6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$ possibilités.

Nombre de Possibilités

1 vers $\sigma_1 \in \{1,2,3,4,5,6\}$.

2 vers $\sigma_2 \in \{1,\ldots,6\} - \{\sigma_1\}$.

3 vers $\sigma_3 \in \{1,\ldots,6\} - \{\sigma_1,\sigma_2\}$.

4 vers $\sigma_4 \in \{1,\ldots,6\} - \{\sigma_1,\ldots,\sigma_3\}$.

5 vers $\sigma_5 \in \{1,\ldots,6\} - \{\sigma_1,\ldots,\sigma_4\}$.

6 vers $\sigma_6 \in \{1,\ldots,6\} - \{\sigma_1,\ldots,\sigma_5\}$.

Total : $6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$ possibilités.

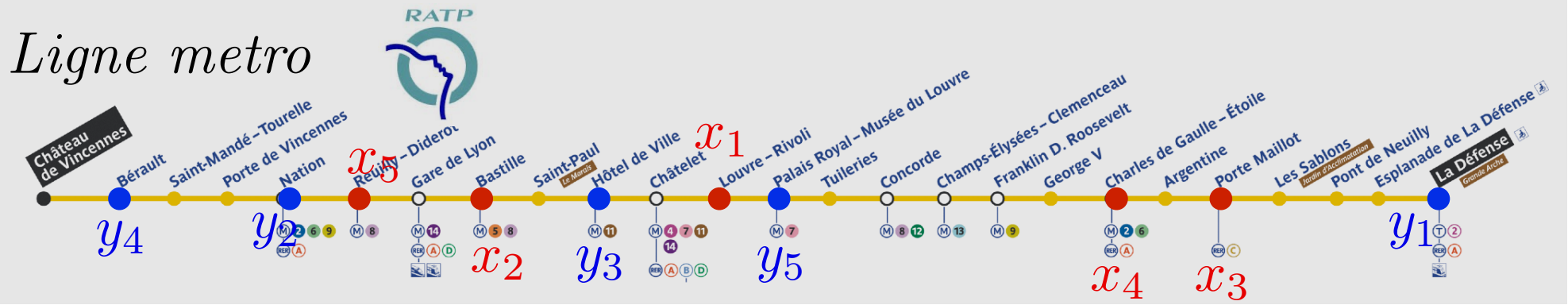
n	*n* !
0	1
1	1
2	2
3	6
4	24
5	120
6	720
7	5 040
8	40 320

n	*n* !
9	362880
10	3628800
11	39916800
12	479001600
…	…
25	$1{,}551 \times 10^{25}$
…	…
70	$1{,}198 \times 10^{100}$

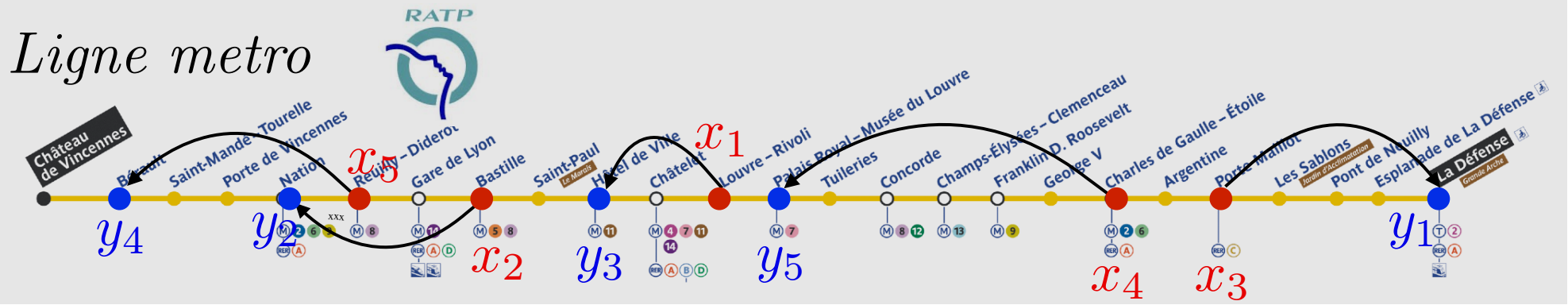
Atomes dans l'univers: 10^{79}

Neurones dans le cerveau: 10^{11}.

En métro

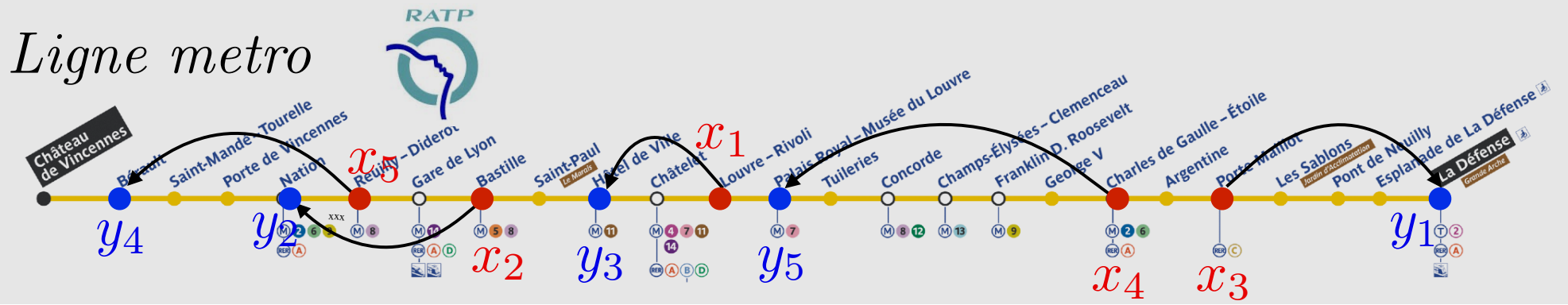

Ligne metro
RATP
Château de Vincennes
Bérault
Saint-Mandé – Tourelle
Porte de Vincennes
Nation
Reuilly – Diderot
Gare de Lyon
Bastille
Saint-Paul
Le Marais
Hôtel de Ville
Châtelet
Louvre – Rivoli
Palais Royal – Musée du Louvre
Tuileries
Concorde
Champs-Élysées – Clemenceau
Franklin D. Roosevelt
George V
Charles de Gaulle – Étoile
Argentine
Porte Maillot
Les Sablons
Jardin d'Acclimatation
Pont de Neuilly
Esplanade de La Défense
La Défense
Grande Arche
y_4
y_2
x_5
x_2
y_3
x_1
y_5
x_4
x_3
y_1

En métro



$\rightarrow$ *Solution:* classer les $(x_i)_i, (y_j)_j$ par ordre croissant.

En métro



$\rightarrow$ *Solution:* classer les $(x_i)_i, (y_j)_j$ par ordre croissant.

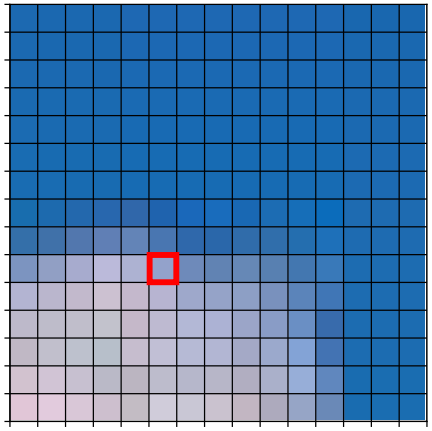
Algorithmes de tri: par selection, pire cas $n(n-1)/2$ opérations.

n	*n* !	n(n-1)/2	n log(n)
10	3628800	45	23
11	39916800	55	26
12	479001600	66	30
25	$1{,}551 \times 10^{25}$	300	80
70	$1{,}198 \times 10^{100}$	21415	297

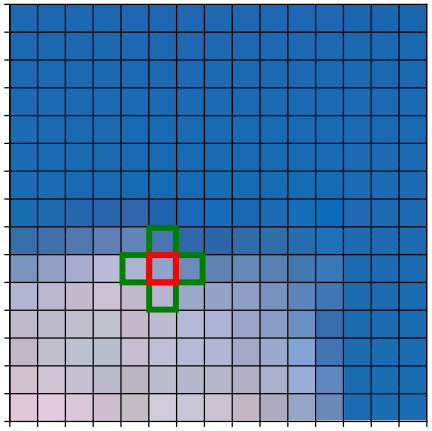
Tri rapide: $n \log(n)$.

Transport optimal en traitement d'image

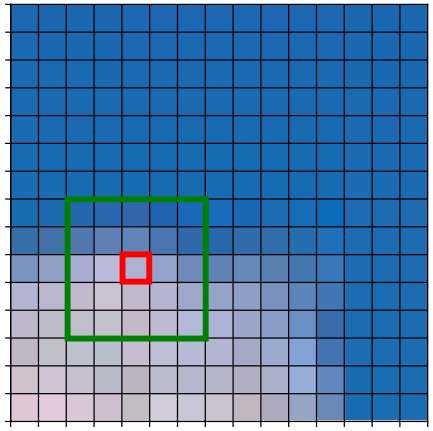
Pixel (couleur)



Pixel+voisins (gradient)



Pixel+zone (style)



Principe: utilisation du transport optimal pour modifier des images

- Image source : image que l'on veut adapter (préservation du contenu sémantique)
- Image cible : exemple de couleur/style que l'on veut donner à l'image source.
- Transformation de Monge pour adapter les propriétés des images.

Applications (du plus simple au plus complexe)

1. Adaptation des couleurs [Ferradans et al., 2014].
2. Transfer de style [Mroueh, 2019].

Image vue comme une distribution en 3D

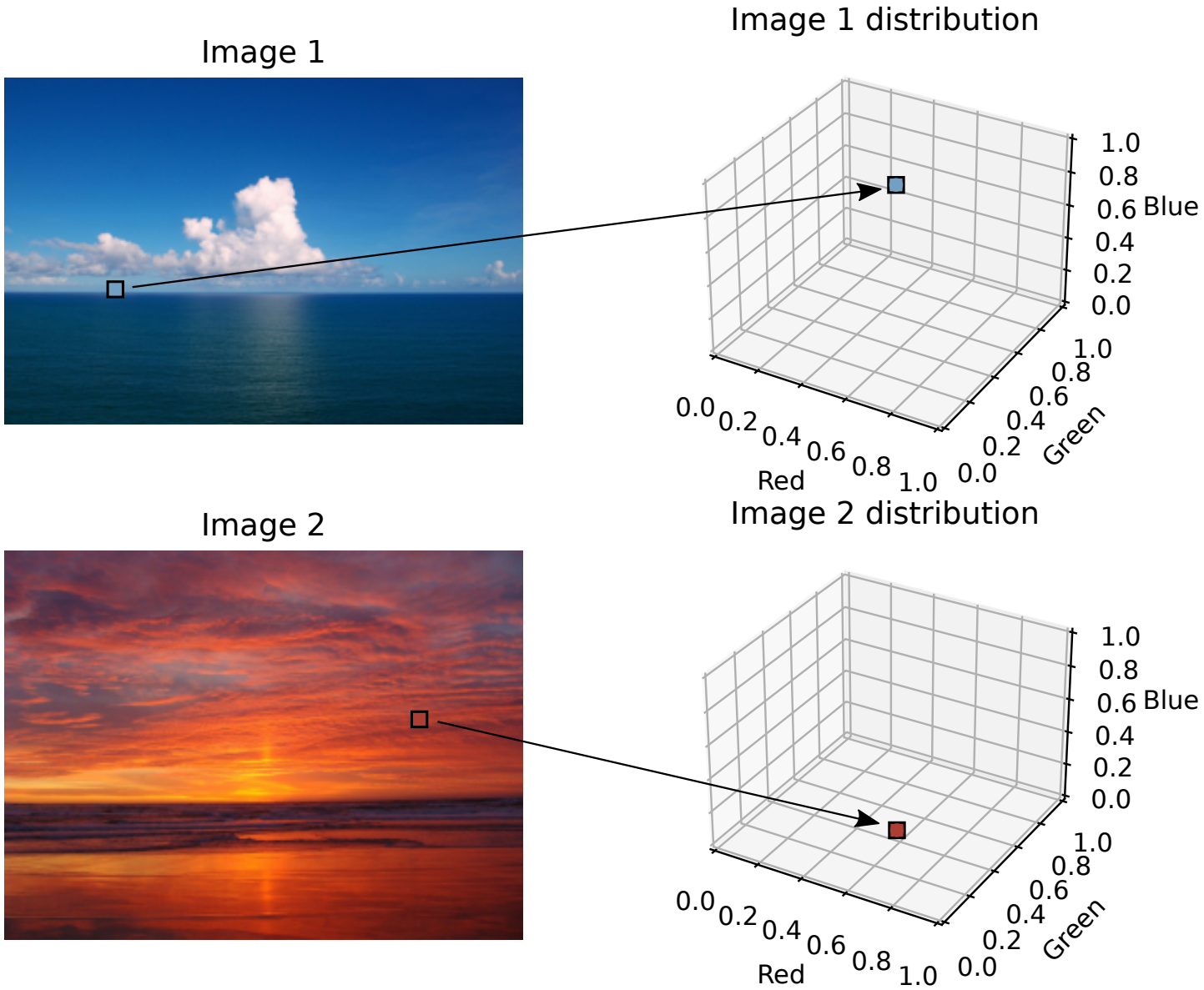


Image vue comme une distribution en 3D

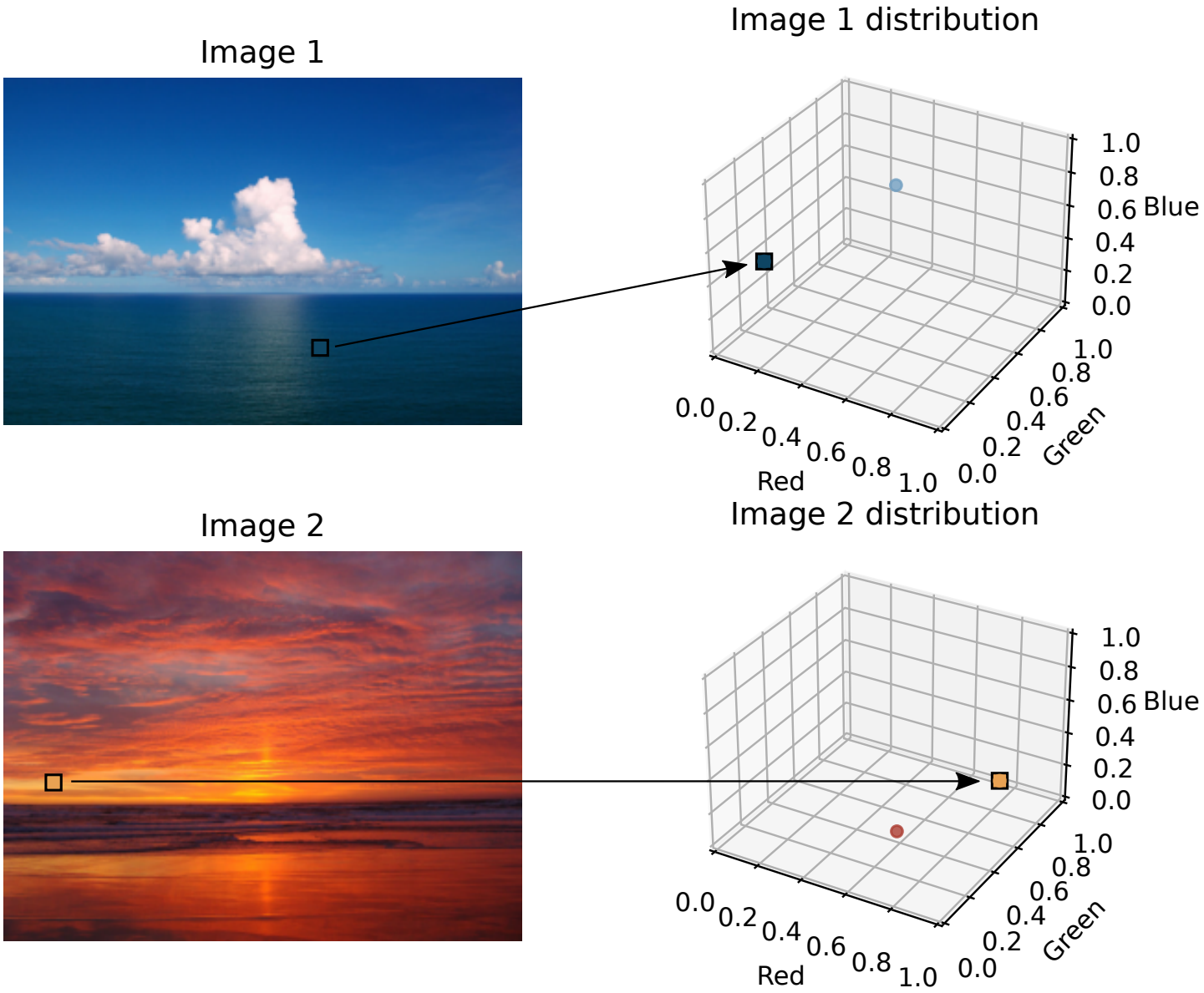


Image vue comme une distribution en 3D

Image 1

Image 1 distribution

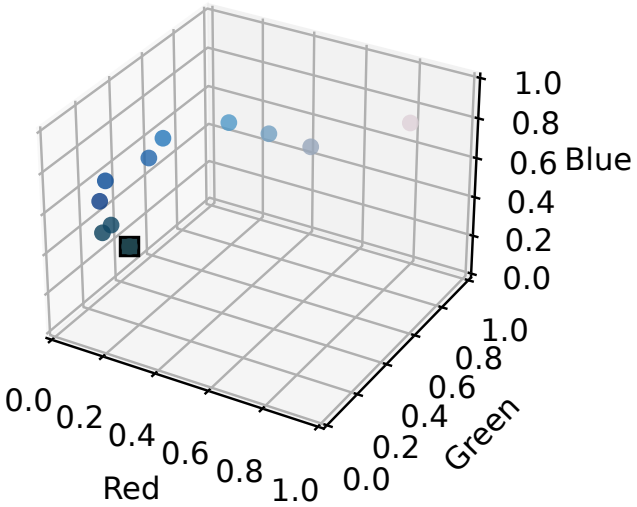


Image 2

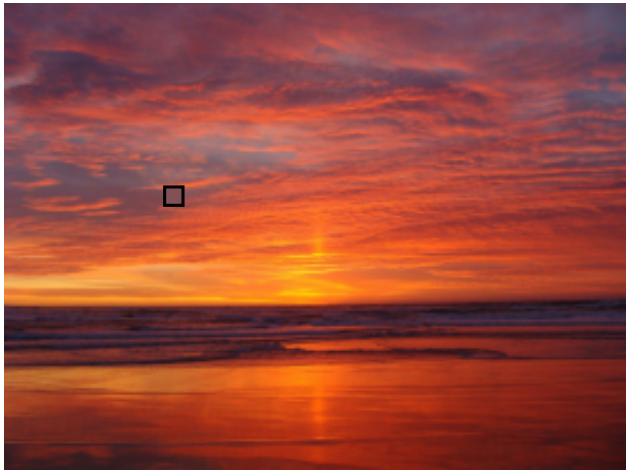


Image 2 distribution

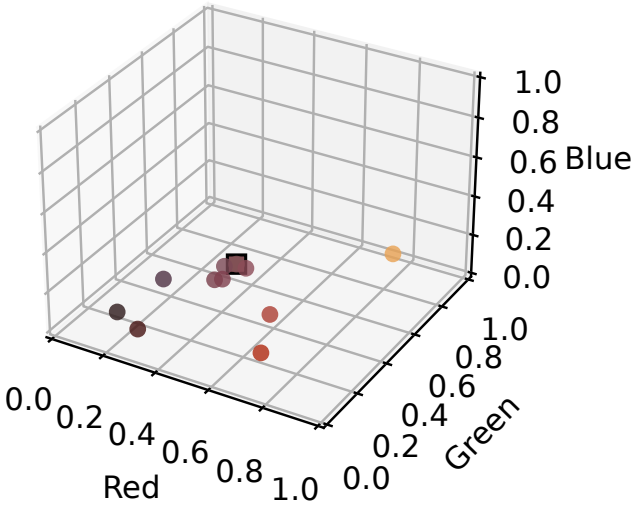


Image vue comme une distribution en 3D

Image 1

Image 1 distribution

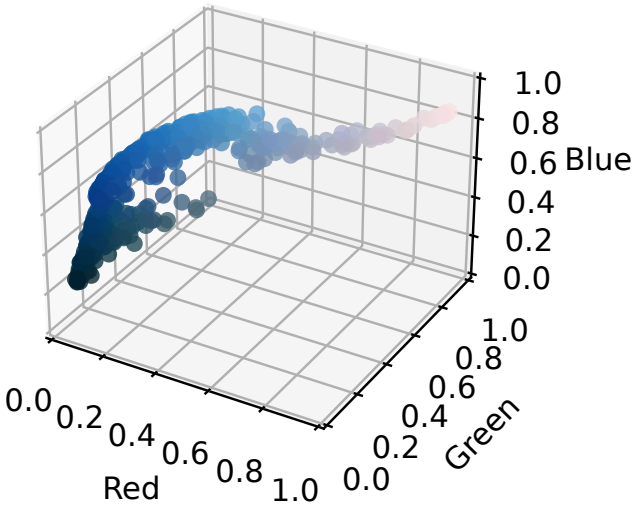
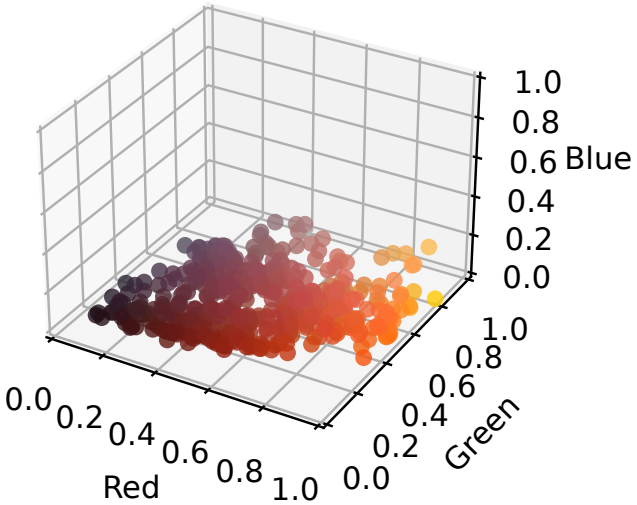


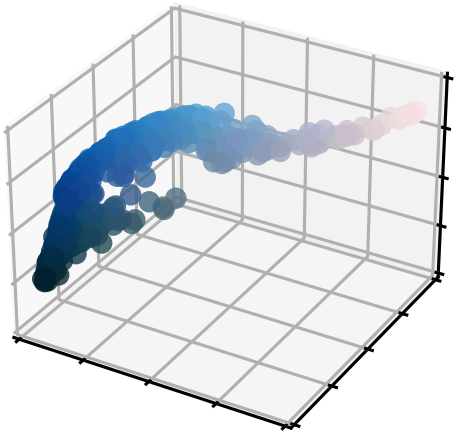
Image 2

Image 2 distribution



Adaptation de couleurs: TO entre pixels

Distribution 1



Distribution 2

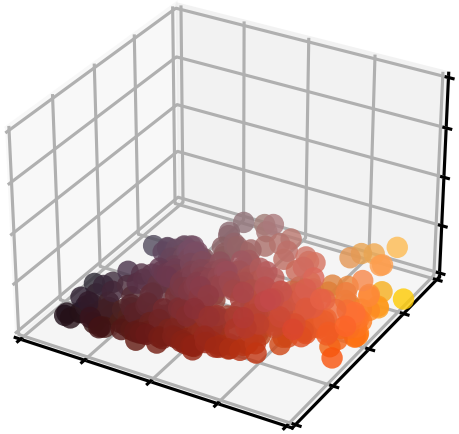
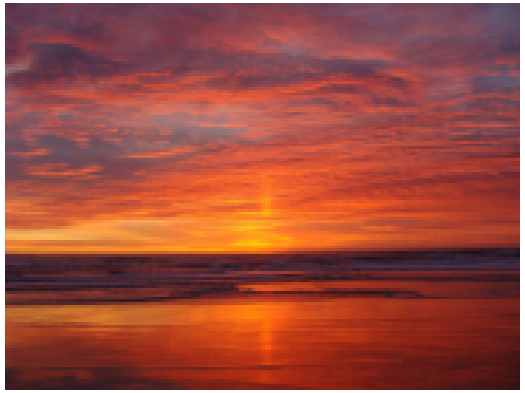


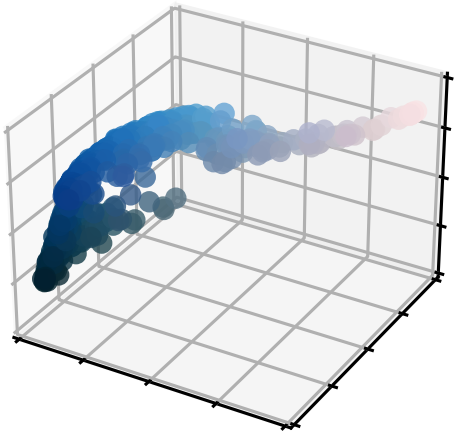
Image 1

Image 2

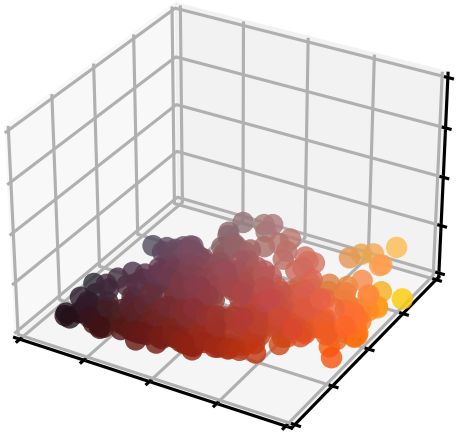


Adaptation de couleurs: TO entre pixels

Distribution 1



Distribution 1 adapted



Distribution 2

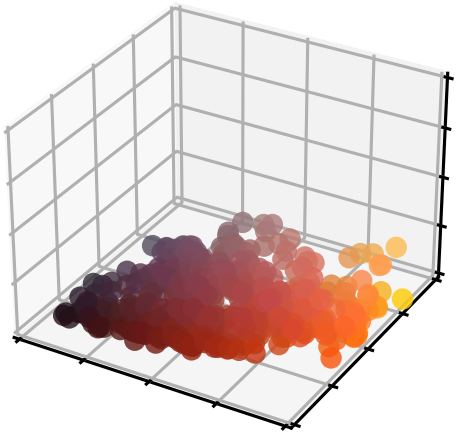
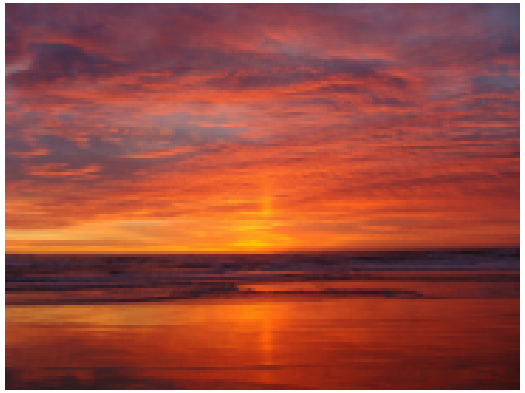


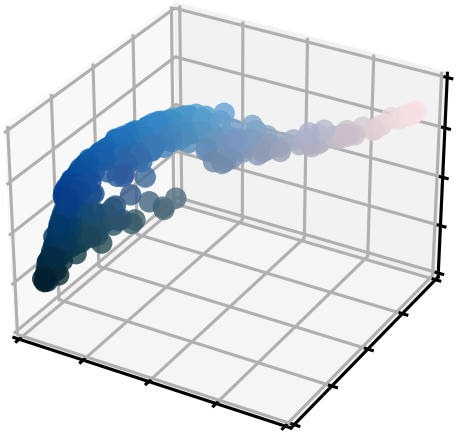
Image 1

Image 2

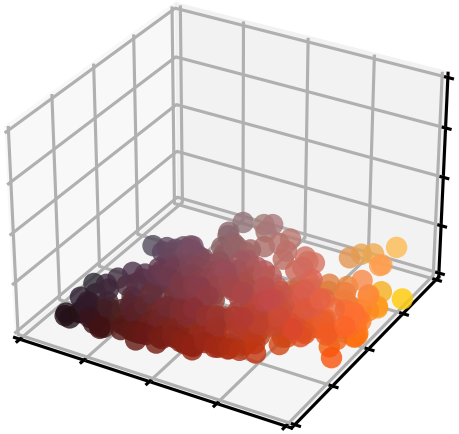


Adaptation de couleurs: TO entre pixels

Distribution 1



Distribution 1 adapted



Distribution 2

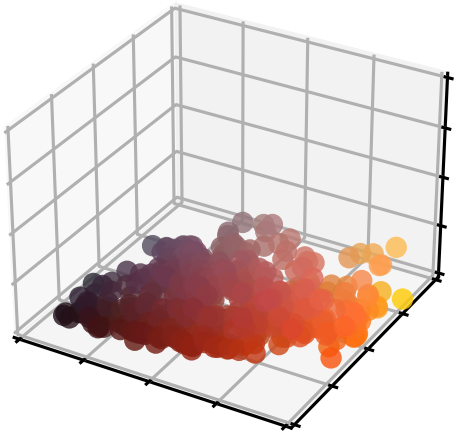


Image 1

Image 1 adapted

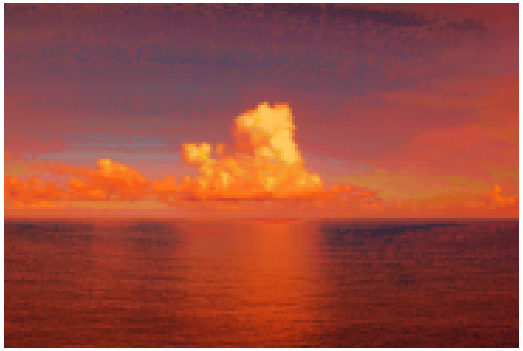
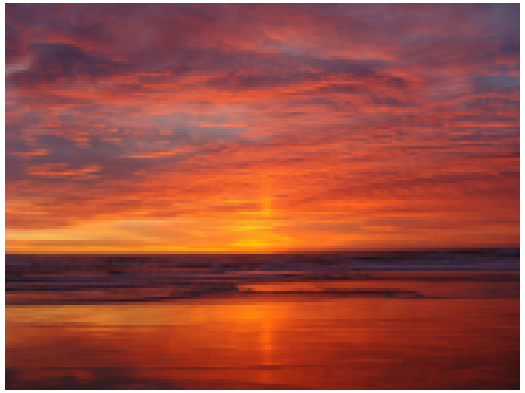
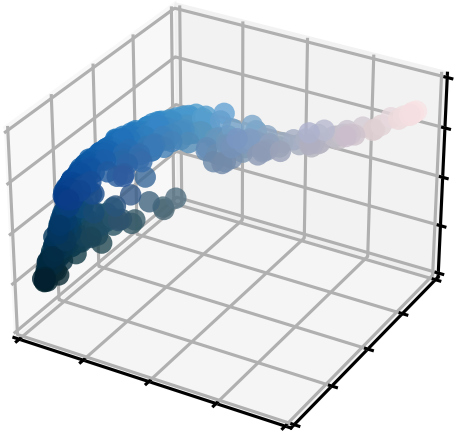


Image 2

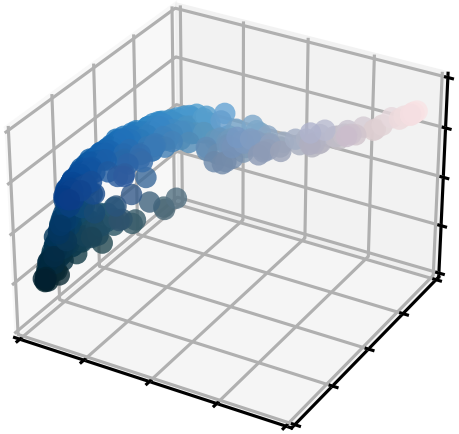


Adaptation de couleurs: TO entre pixels

Distribution 1



Distribution 2 adapted



Distribution 2

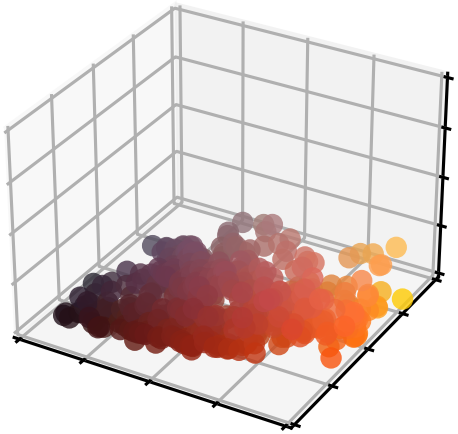
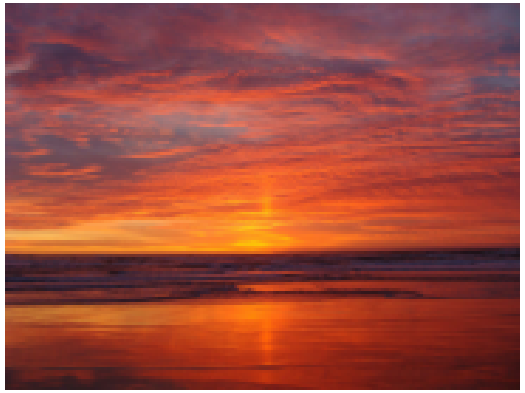


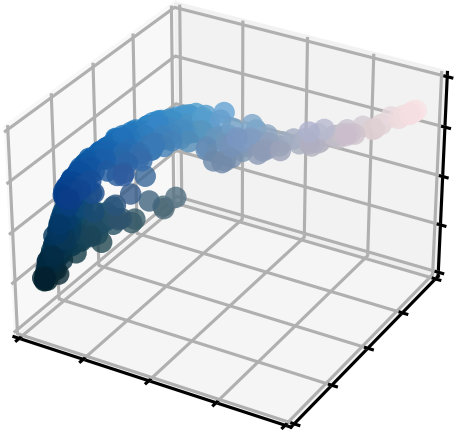
Image 1

Image 2

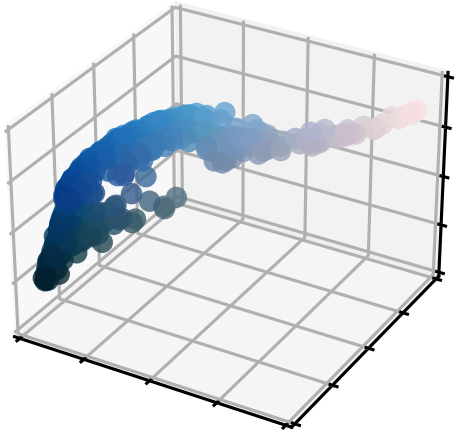


Adaptation de couleurs: TO entre pixels

Distribution 1



Distribution 2 adapted



Distribution 2

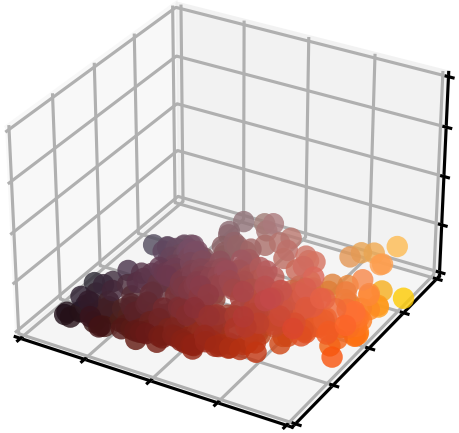


Image 1

Image 2 adapted

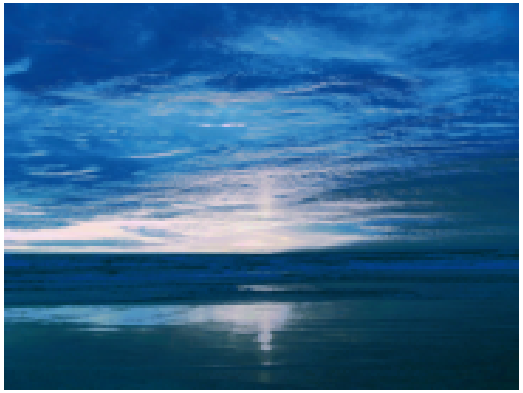
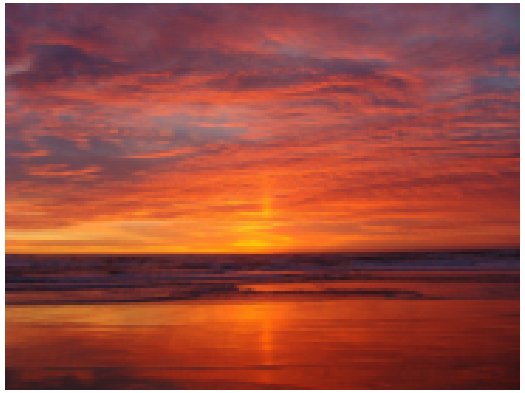
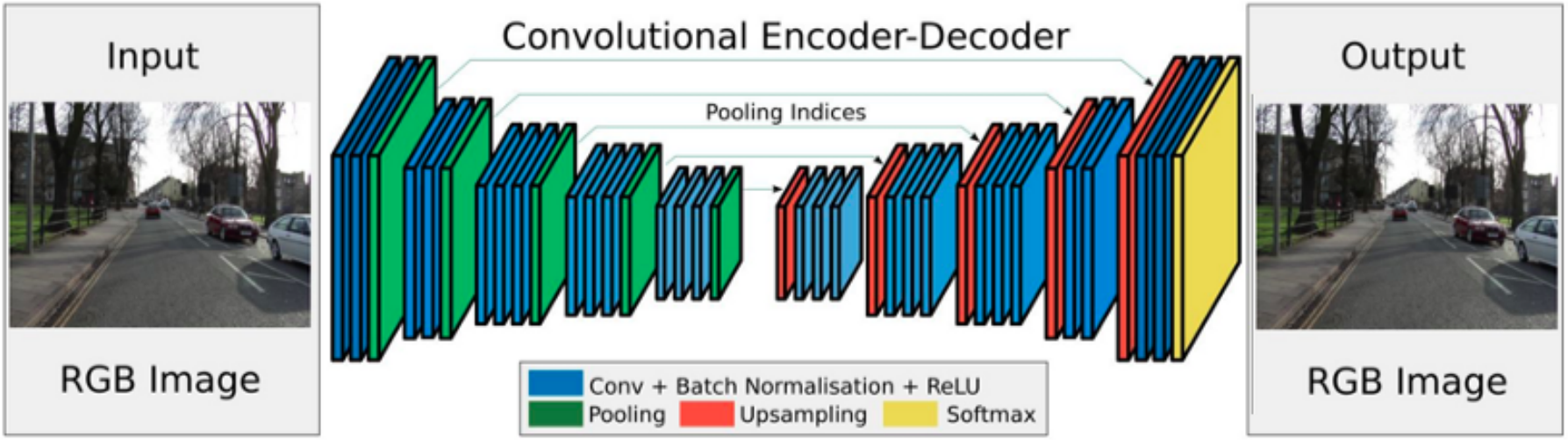


Image 2



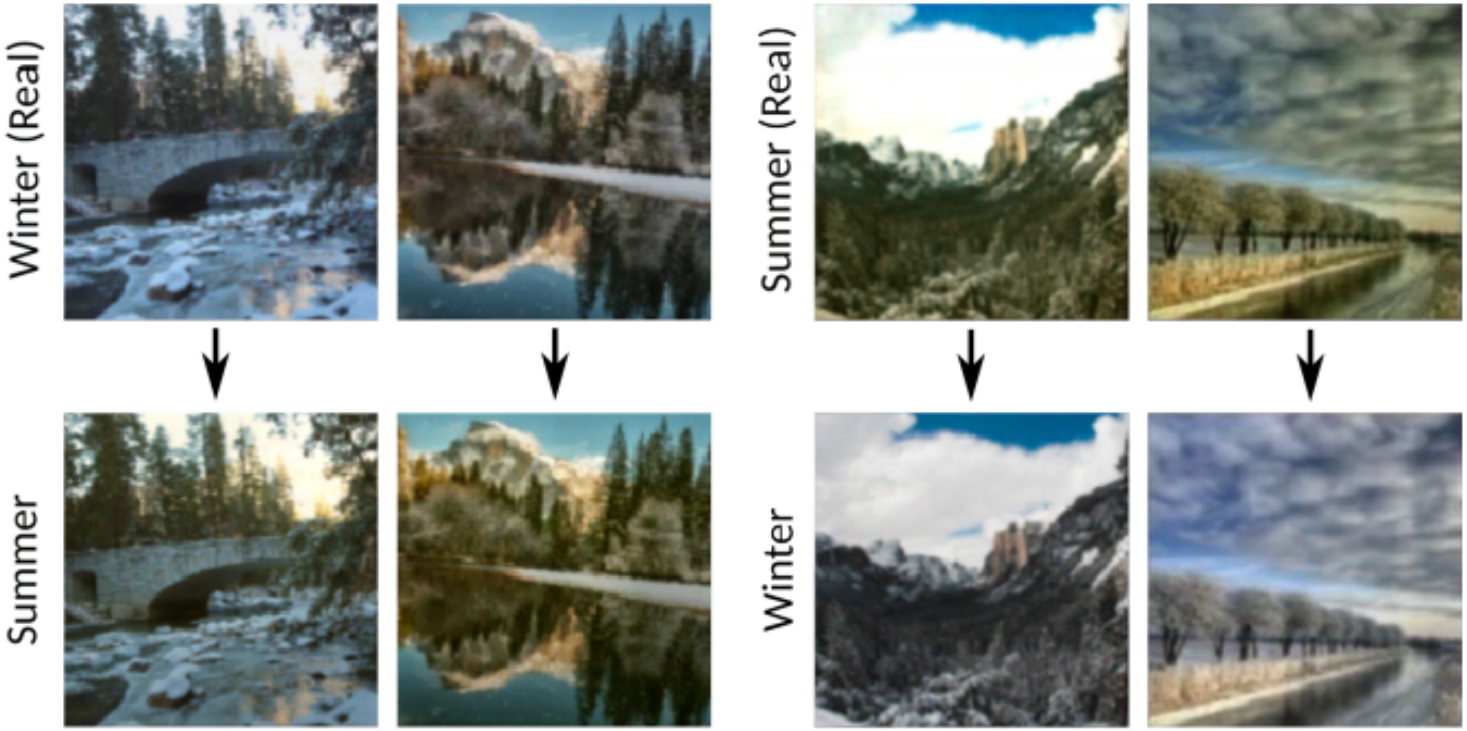
Représentation d'image avec modèles profonds



Principe de l'autoencodeur

- Réseau de neurone visant à apprendre une représentation compacte d'un ensemble de données (images).
- Les images sont encodées par patchs (zones) au lieu de pixels ce qui permet de modéliser les textures.
- Les distributions de patchs peuvent être modifiées (ce qui modifie l'image).

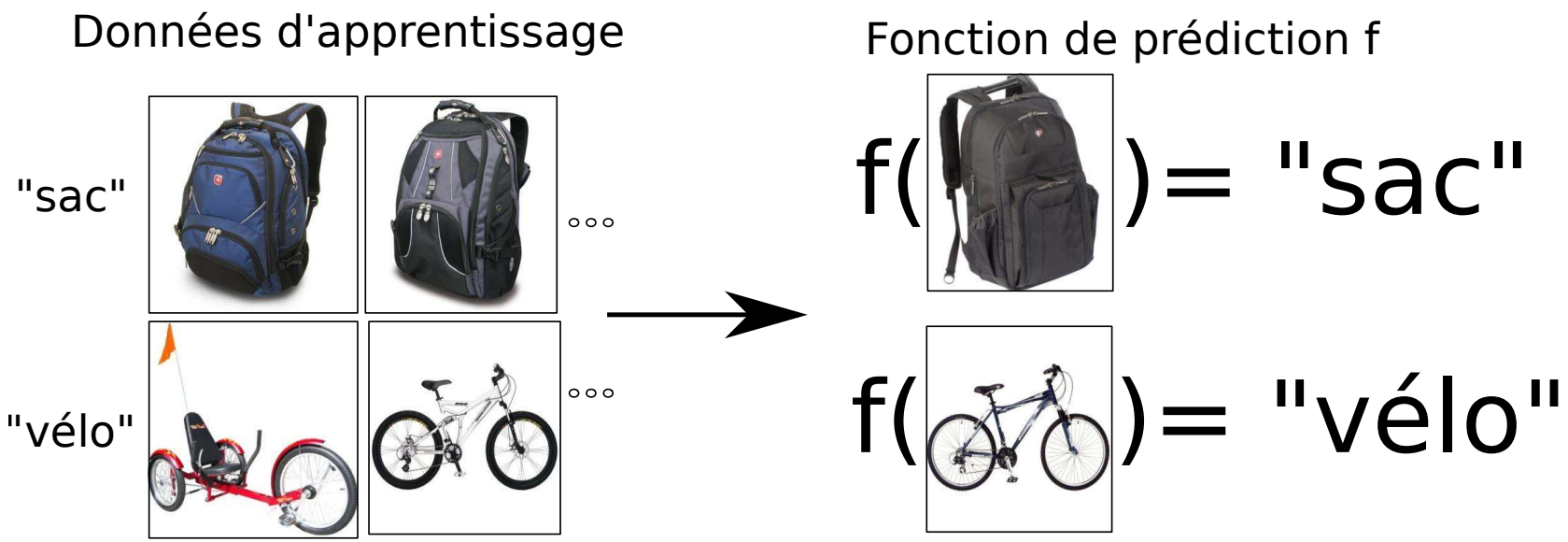
Transport optimal pour le transfer de style



Principe

- Utiliser un autoencodeur pour encoder l'image et extraire les distributions de la source et cible.
- Transporter les distributions des patchs dans entre image source et cible puis reconstruire la source transportée.
- Transformation de Monge linéaire (Wasserstein Style Transfer [Mroueh, 2019]) non linéaire [Korotin et al., 2019].

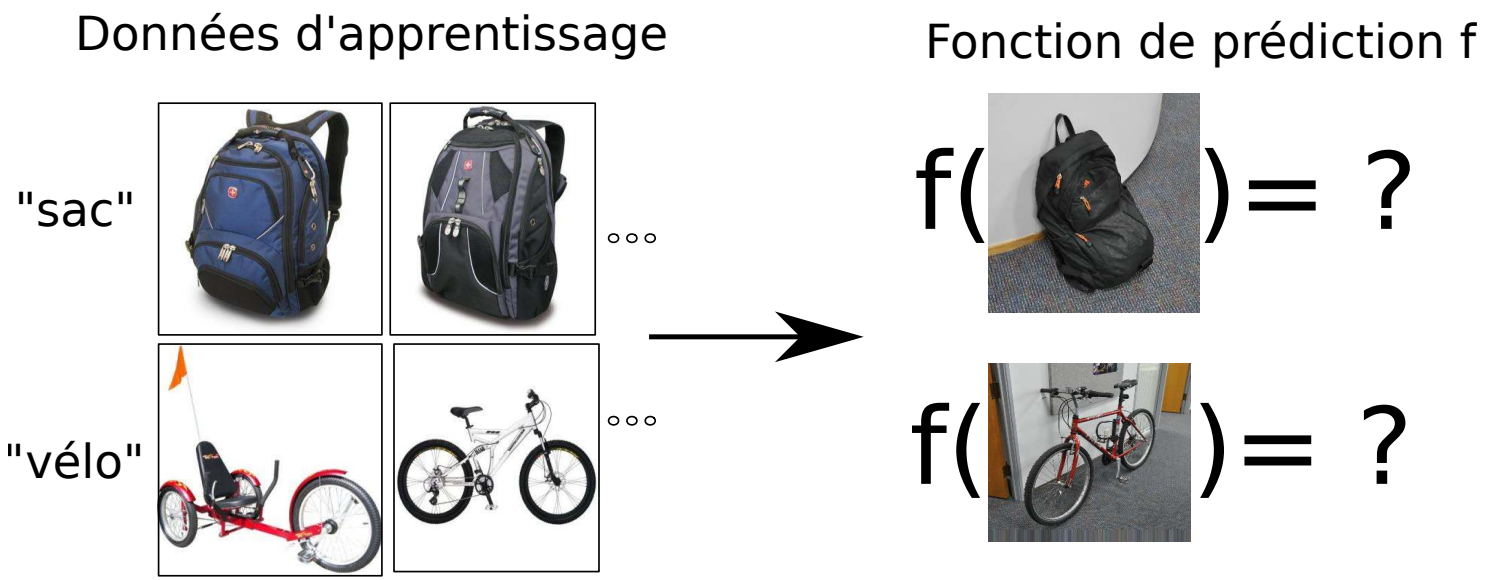
Classification supervisée et décalage de données



Contexte (data shift)

- Apprentissage de classification automatique à partir de données d'apprentissage.
- Décalage de données: les données de test sont différentes des données d'entraînement, perte de performance.
- Exemple : changement de capteur, changement de conditions météorologiques, changement de période de l'année, simulation VS vraies données.

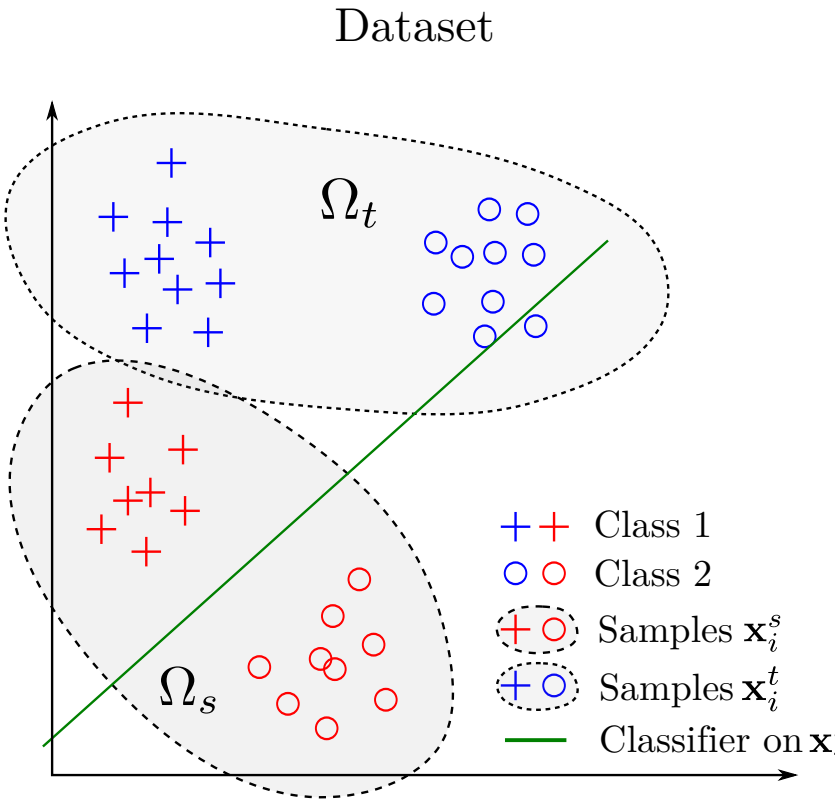
Classification supervisée et décalage de données



Contexte (data shift)

- Apprentissage de classification automatique à partir de données d'apprentissage.
- Décalage de données: les données de test sont différentes des données d'entraînement, perte de performance.
- Exemple : changement de capteur, changement de conditions météorologiques, changement de période de l'année, simulation VS vraies données.

Classification supervisée et décalage de données



Contexte (data shift)

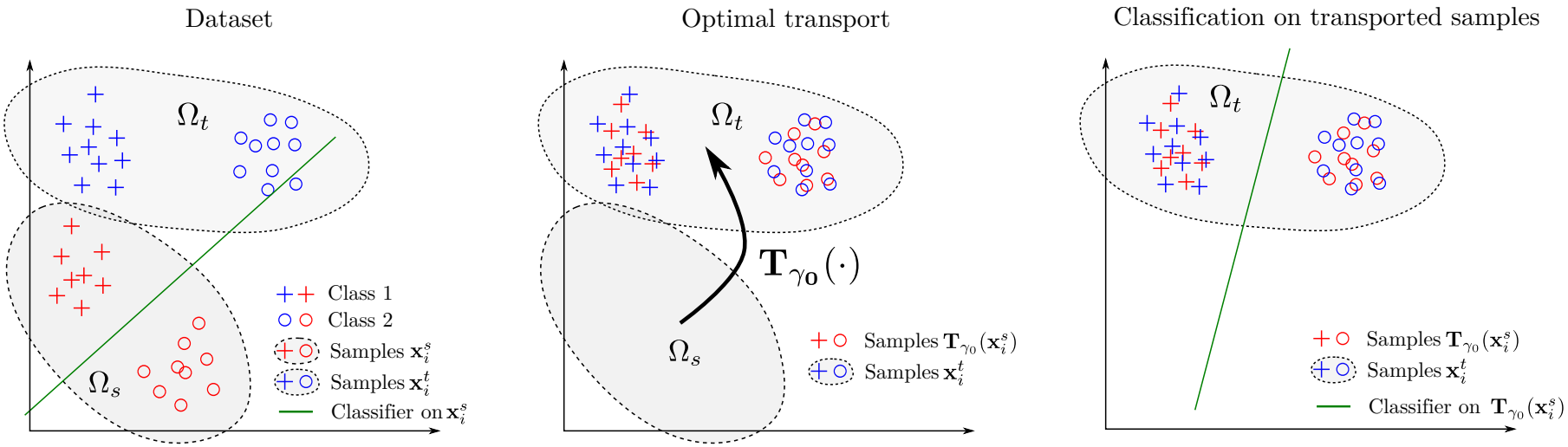
- Apprentissage de classification automatique à partir de données d'apprentissage.
- Décalage de données: les données de test sont différentes des données d'entraînement, perte de performance.
- Exemple : changement de capteur, changement de conditions météorologiques, changement de période de l'année, simulation VS vraies données.

Le problème de l'adaptation de domaine

Données et objectif

- Les classes des examples sont disponible dans le jeux de données source mais pas de classes sont disponible en cible.
- L'objectif de l'adaptation de domaine ets d'apprendre un prédicteur qui prédit sur les données cibles en n'ayant que les classes sur données source.

Transport Optimal pour l'adaptation de domaine



Principe [Courty et al., 2016]

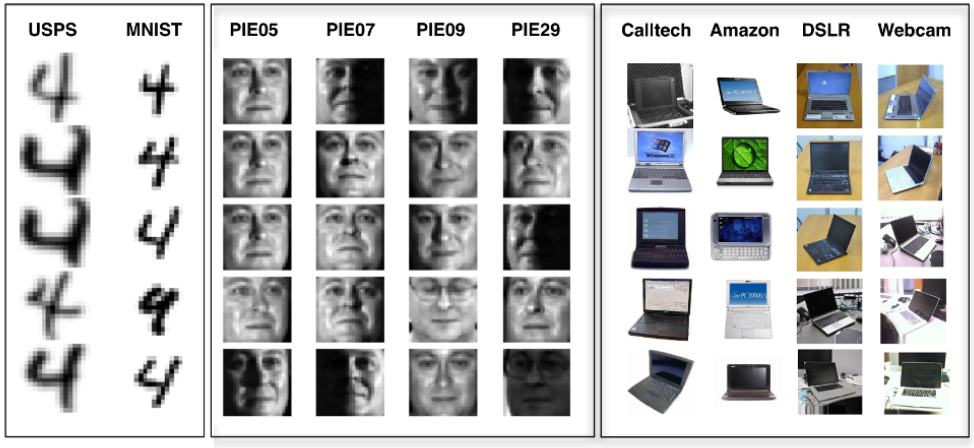
1. Estimation d'une transformation de Monge entre les données source et cible.
2. Transport des données source vers cible (elle conservent leur classe)
3. Apprentissage d'un classifieur sur les données transportées.

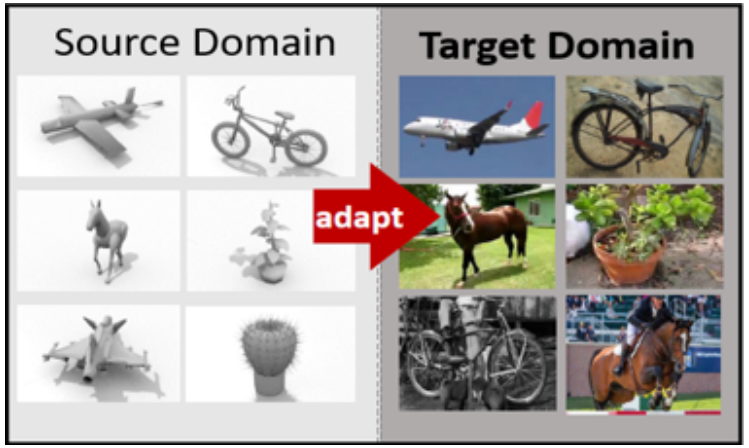
Intuition

- Si les données ont changé entre les source et cible, on peut compenser ce changement en estimant ce qu'il s'ets passé.
- Parmi tous les transformations possibles, on prend la plus simple, c'est-à-dire celui du moindre effort: Transport Optimal.

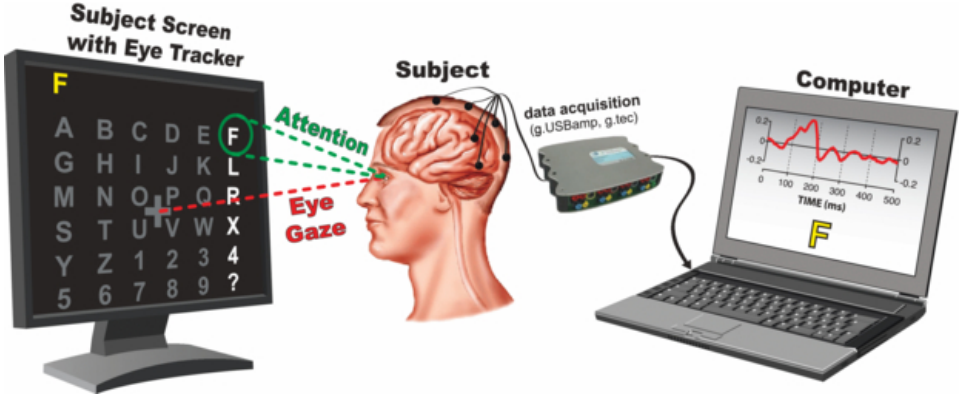
Applications en adaptation de domaine

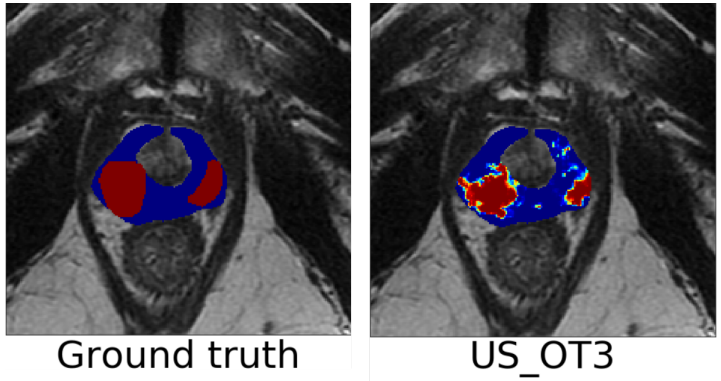
Vision par ordinateur





Traitement de données biomédicales



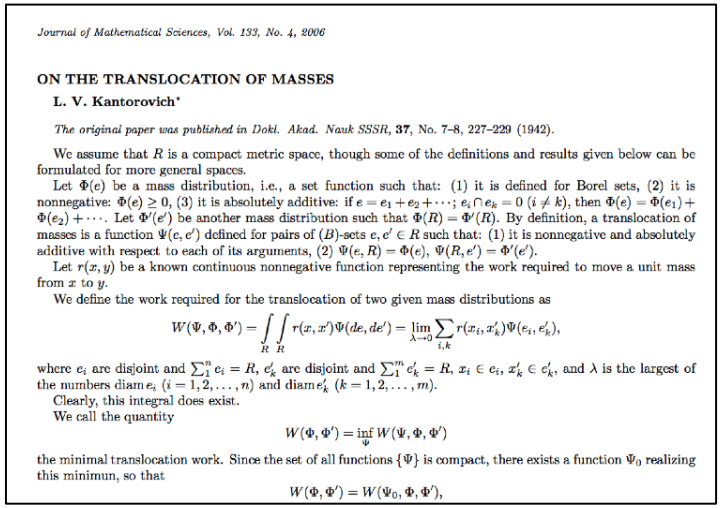


Leonid Kantorovich (1912-1986)

Леонид Витальевич Канторович

Au cours du siège de Léningrad, Kantorovitch est responsable de la sécurité de la Route de la vie. Il détermine la distance optimale à observer entre les voitures sur la surface gelée du lac Ladoga, en fonction de l'épaisseur de glace et de la température de l'air. En décembre 1941–janvier 1942, Kantorovitch s'assure lui-même de la viabilité de la banquise en marchant entre les camions. Cependant bien des véhicules chargés d'approvisionnements sont détruits par les bombardements aériens nazis. En récompense de ses exploits et de son courage, les autorités attribuent à Kantorovitch l'ordre de la Guerre patriotique, et le décorent de la Médaille pour la Défense de Léningrad.

Peu avant la Seconde Guerre mondiale, Leonid Kantorovitch découvre la programmation linéaire, l'optimisation linéaire et ses applications à l'optimisation de la production économique planifiée. Il est le seul chercheur soviétique à avoir reçu le « prix Nobel » d'économie (1975). Les théories de Kantorovitch ne sont publiées qu'après l'ère stalinienne. Un des apports de Kantorovitch est d'avoir incité à une meilleure prise en compte de la productivité marginale de l'investissement, afin de résoudre les difficultés liées à l'allocation des ressources au sein d'une économie socialiste.



Journal of Mathematical Sciences, Vol. 133, No. 4, 2006

ON THE TRANSLOCATION OF MASSES

L. V. Kantorovich*

The original paper was published in Dokl. Akad. Nauk SSSR, **37**, No. 7–8, 227–229 (1942).

We assume that R is a compact metric space, though some of the definitions and results given below can be formulated for more general spaces.

Let $\Phi(e)$ be a mass distribution, i.e., a set function such that: (1) it is defined for Borel sets, (2) it is nonnegative: $\Phi(e) \geq 0$, (3) it is absolutely additive: if $e = e_1 + e_2 + \cdots$; $e_i \cap e_k = 0$ $(i \neq k)$, then $\Phi(e) = \Phi(e_1) + \Phi(e_2) + \cdots$. Let $\Phi'(e')$ be another mass distribution such that $\Phi(R) = \Phi'(R)$. By definition, a translocation of masses is a function $\Psi(e, e')$ defined for pairs of (B)-sets $e, e' \in R$ such that: (1) it is nonnegative and absolutely additive with respect to each of its arguments, (2) $\Psi(e, R) = \Phi(e)$, $\Psi(R, e') = \Phi'(e')$.

Let $r(x, y)$ be a known continuous nonnegative function representing the work required to move a unit mass from x to y.

We define the work required for the translocation of two given mass distributions as

$$W(\Psi, \Phi, \Phi') = \int_R \int_R r(x, x')\Psi(de, de') = \lim_{\lambda \to 0} \sum_{i,k} r(x_i, x'_k)\Psi(e_i, e'_k),$$

where e_i are disjoint and $\sum_1^n e_i = R$, e'_k are disjoint and $\sum_1^m e'_k = R$, $x_i \in e_i$, $x'_k \in e'_k$, and λ is the largest of the numbers $\operatorname{diam} e_i$ $(i = 1, 2, \ldots, n)$ and $\operatorname{diam} e'_k$ $(k = 1, 2, \ldots, m)$.

Clearly, this integral does exist.

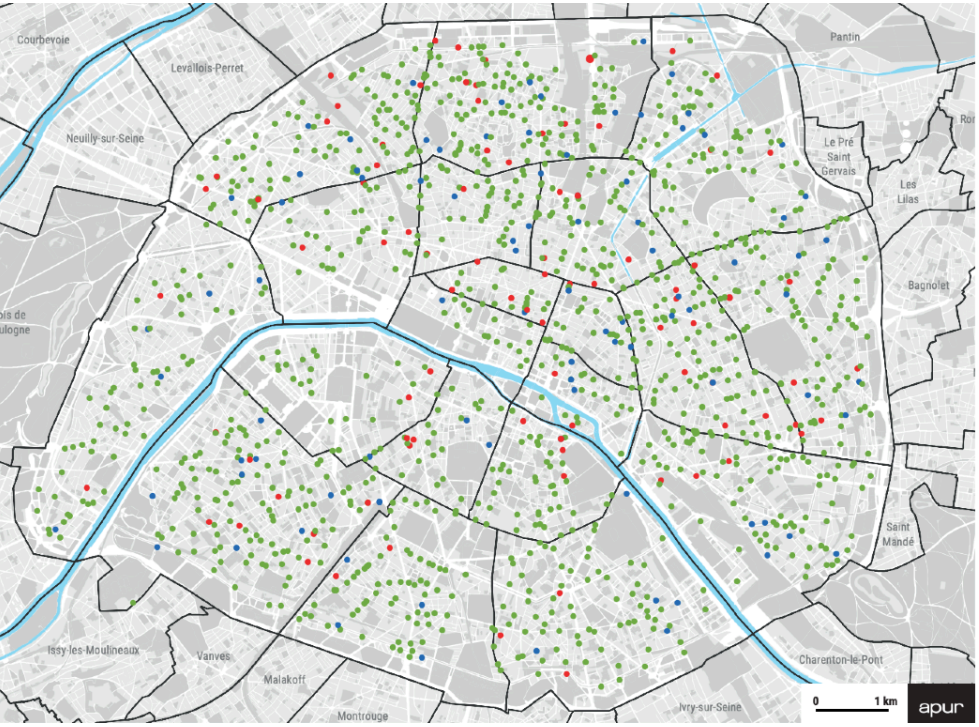
We call the quantity

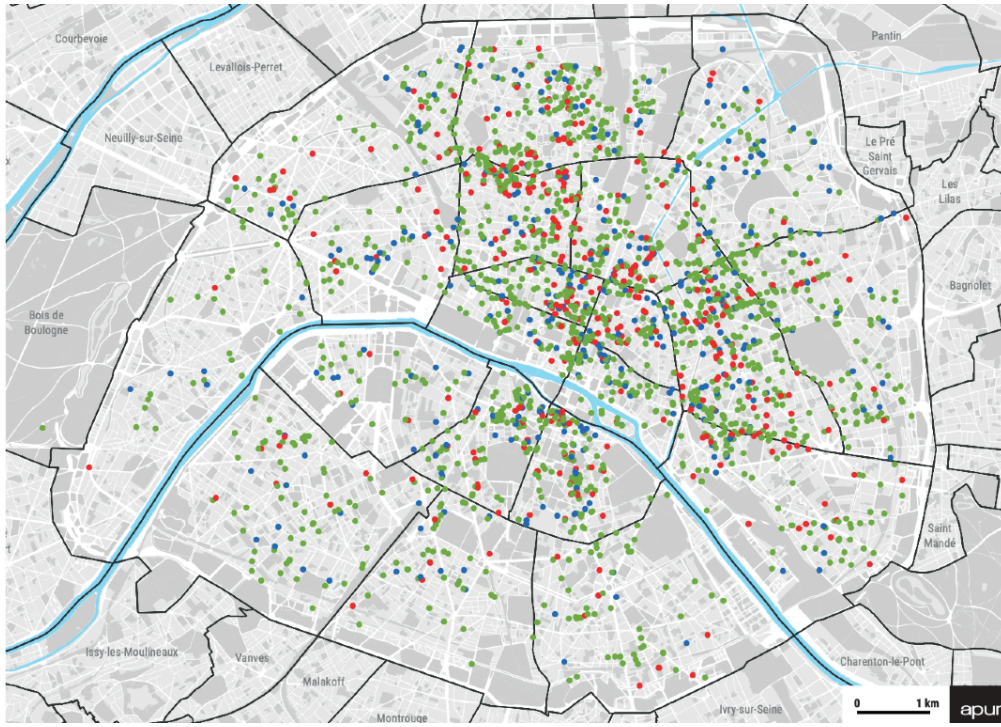
$$W(\Phi, \Phi') = \inf_{\Psi} W(\Psi, \Phi, \Phi')$$

the minimal translocation work. Since the set of all functions $\{\Psi\}$ is compact, there exists a function Ψ_0 realizing this minimun, so that

$$W(\Phi, \Phi') = W(\Psi_0, \Phi, \Phi'),$$

Dans Paris





1085 boulangeries < **1410 cafés**

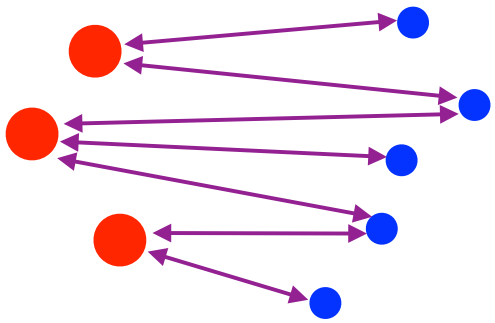
+ bars, etc.

Transport de Kantorovitch

Affectation entre ensembles de tailles $n \neq m$: *impossible*.

Chaque point a une masse $\frac{1}{n}$ et $\frac{1}{m}$.

Autoriser la division de la masse.

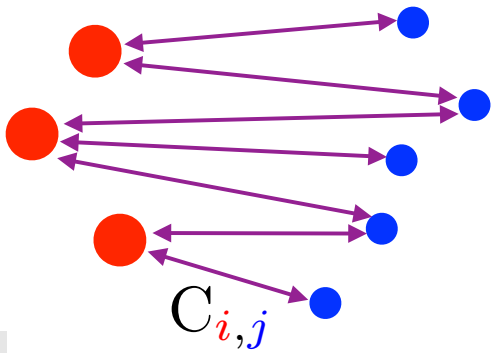


Transport de Kantorovitch

Affectation entre ensembles de tailles $n \neq m$: *impossible*.

Chaque point a une masse $\frac{1}{n}$ et $\frac{1}{m}$.

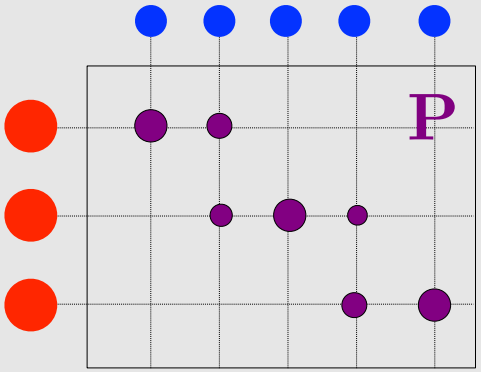
Autoriser la division de la masse.



Plan de transport:

$$\sum_j \mathbf{P}_{i,j} = \frac{1}{n}$$

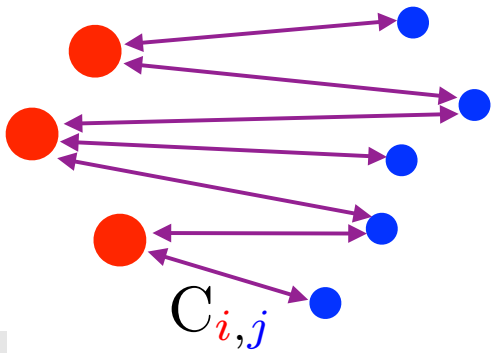
$$\sum_i \mathbf{P}_{i,j} = \frac{1}{m}$$



Transport de Kantorovitch

Affectation entre ensembles de tailles $n \neq m$: *impossible*.

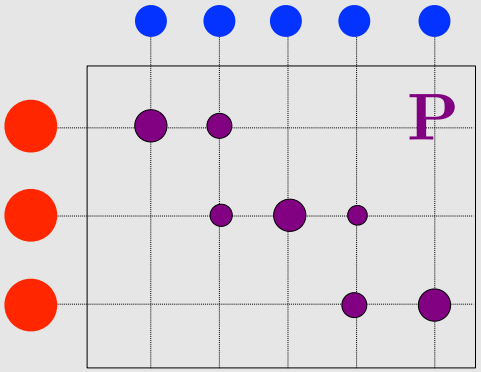
- Chaque point a une masse $\frac{1}{n}$ et $\frac{1}{m}$.
- Autoriser la division de la masse.



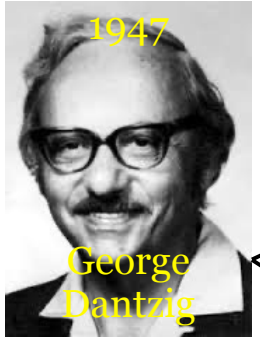
Plan de transport:

$$\sum_j \mathbf{P}_{i,j} = \frac{1}{n}$$

$$\sum_i \mathbf{P}_{i,j} = \frac{1}{m}$$



Transport optimal: $\min_{\mathbf{P} \geqslant 0} \sum_{i,j} \mathrm{C}_{i,j} \mathbf{P}_{i,j}$

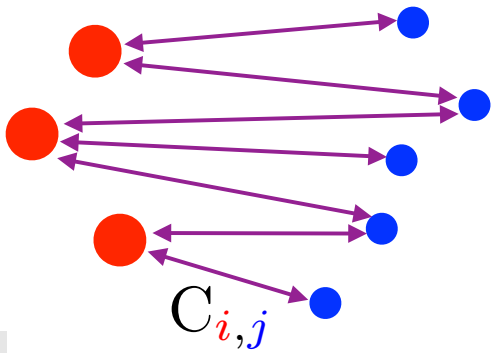


- Problème d'optimisation (programmation linéaire).
- Algorithme du simplex: solution en n^3 operations.

Transport de Kantorovitch

Affectation entre ensembles de tailles $n \neq m$: *impossible*.

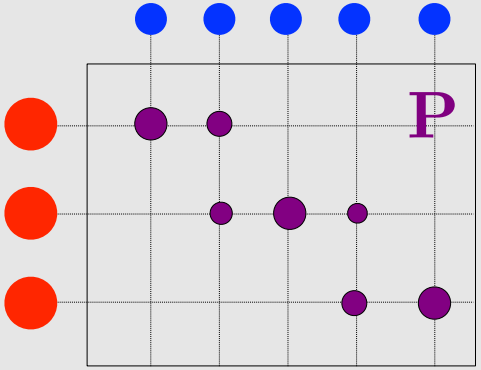
- Chaque point a une masse $\frac{1}{n}$ et $\frac{1}{m}$.
- Autoriser la division de la masse.



Plan de transport:

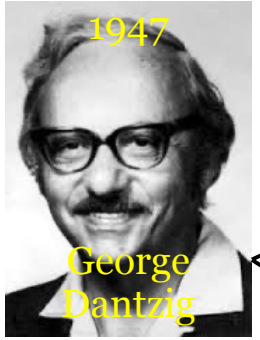
$$\sum_j \mathbf{P}_{i,j} = \frac{1}{n}$$

$$\sum_i \mathbf{P}_{i,j} = \frac{1}{m}$$



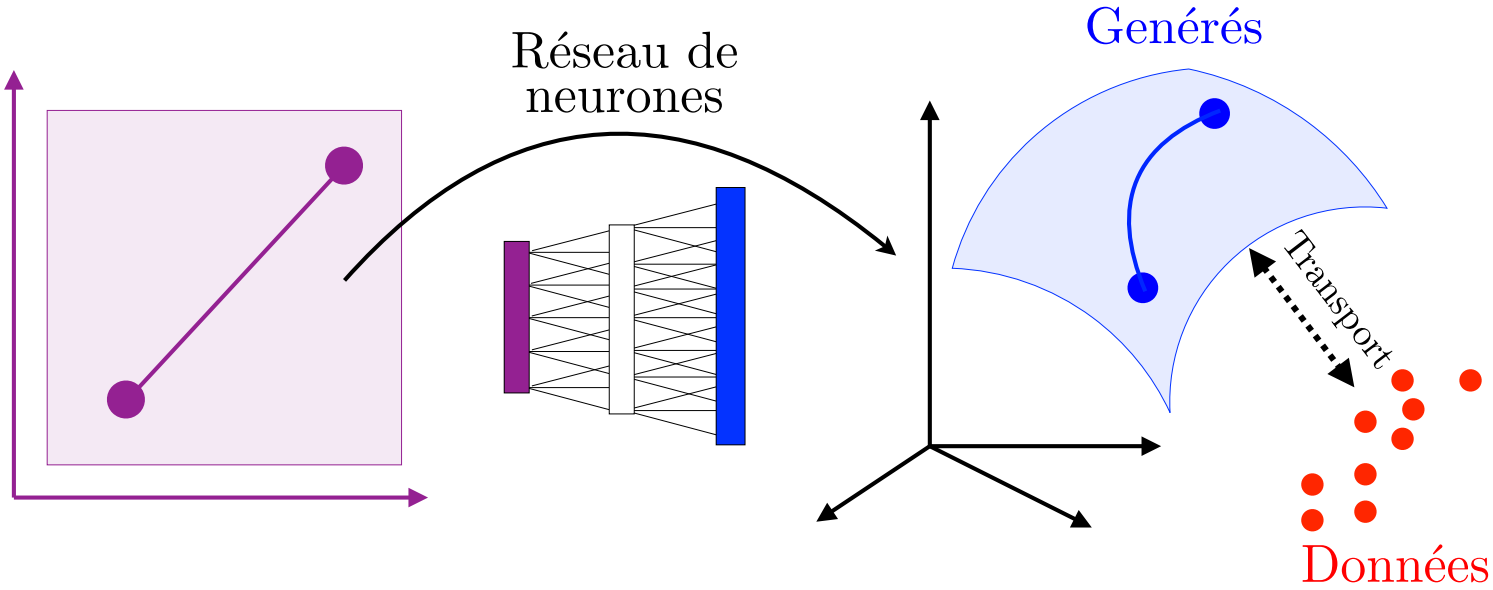
Transport optimal: $\min_{\mathbf{P} \geqslant 0} \sum_{i,j} \mathrm{C}_{i,j} \mathbf{P}_{i,j}$

Distance de transport $=: \mathrm{W}(X, Y)$



- Problème d'optimisation (programmation linéaire).
- Algorithme du simplex: solution en n^3 operations.

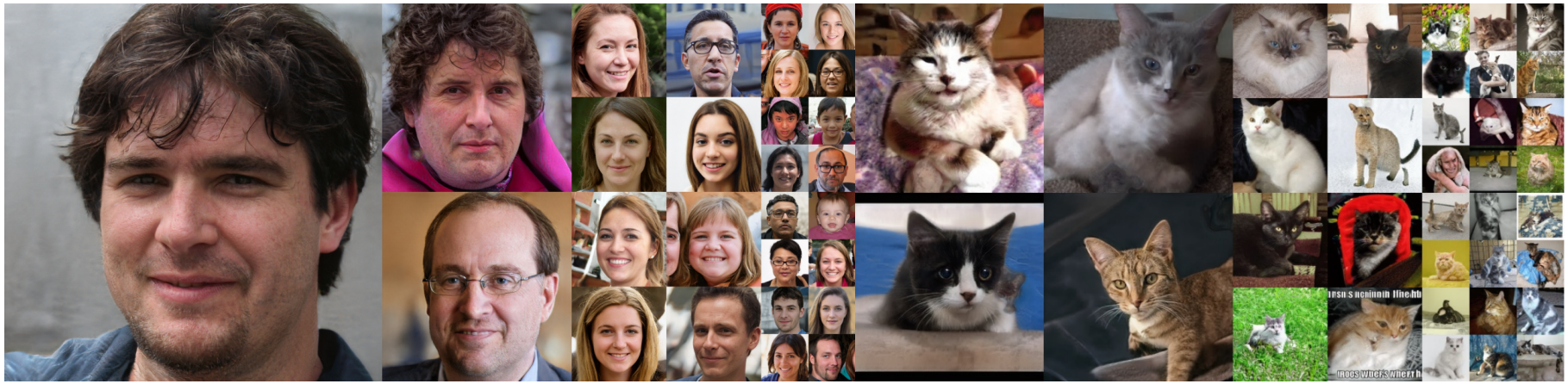
Réseaux de neurones génératifs



Objectif: Générer des données réalistes

- Principe [Goodfellow et al., 2014]
  - Générer un vecteur de bruit z (connu, facile à générer).
  - Trouver G un réseau de neurones tel que $G(z)$ est une image réaliste ($\forall z$).
- Wasserstein GAN [Arjovsky et al., 2017] minimise la distance de Wasserstein entre les données générées et les vraies données.

Réseaux de neurones génératifs

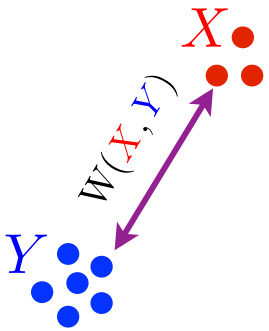


Objectif: Générer des données réalistes

- Principe [Goodfellow et al., 2014]
  - Générer un vecteur de bruit z (connu, facile à générer).
  - Trouver G un réseau de neurones tel que $G(z)$ est une image réaliste ($\forall z$).
- Wasserstein GAN [Arjovsky et al., 2017] minimise la distance de Wasserstein entre les données générées et les vraies données.
- Utilisé dans StyleGAN [Karras et al., 2019] pour générer des visages de personnes particulièrement réalistes (et des chats).

Barycentres de transport optimal

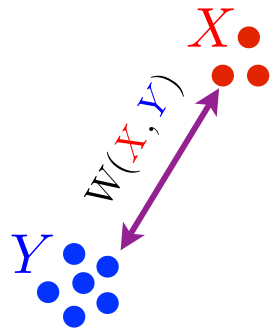
Distance de Wasserstein: $\mathrm{W}(X, Y) = \min_{\mathbf{P} \geqslant 0} \sum_{i,j} \mathrm{C}_{i,j} \mathbf{P}_{i,j}$

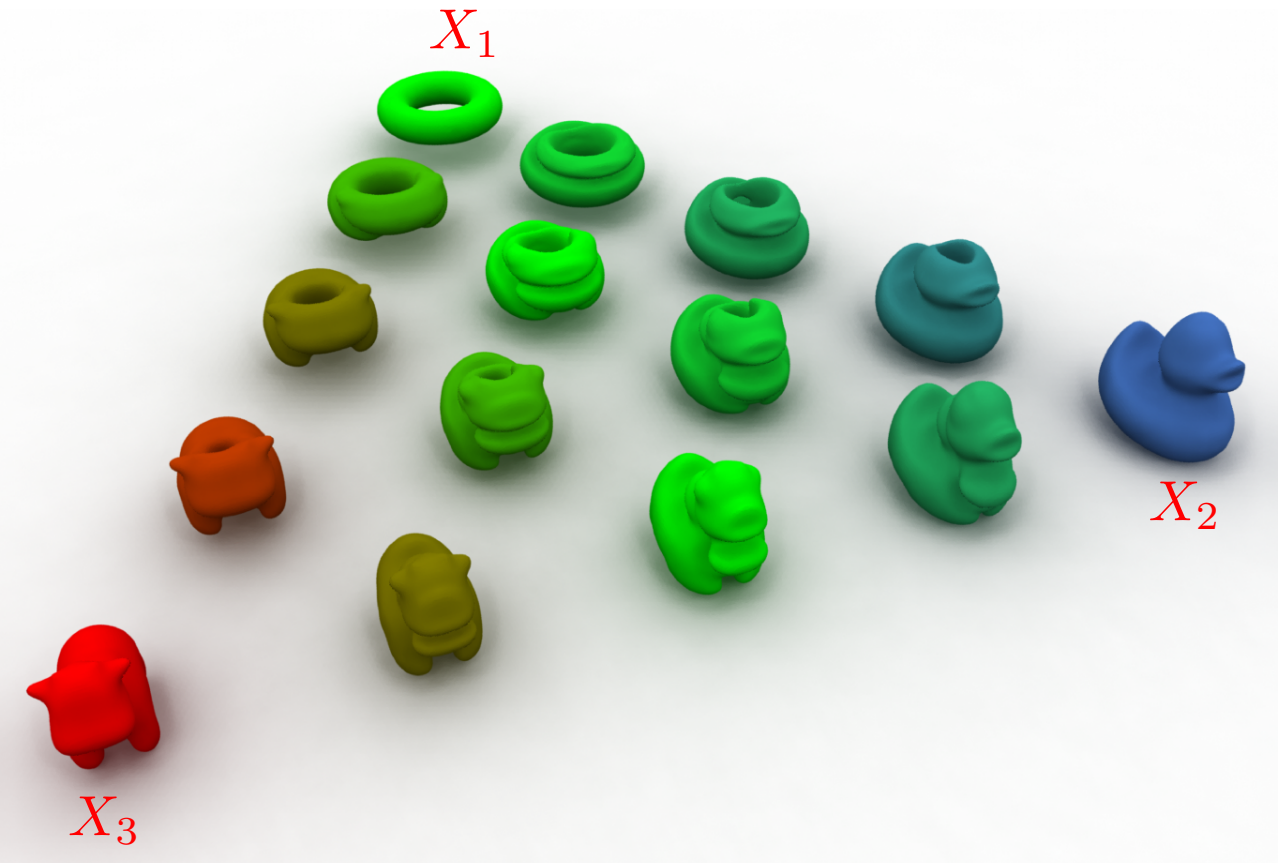


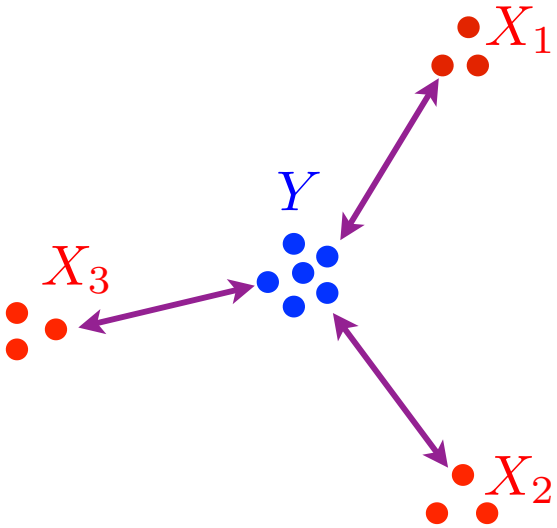
Barycentres de transport optimal

Distance de Wasserstein: $\mathrm{W}(X,Y) = \min_{\mathbf{P}\geqslant 0} \sum_{i,j} \mathrm{C}_{i,j}\mathbf{P}_{i,j}$

Barycenters de Wasserstein: $\min_{Y} \lambda_1 \mathrm{W}(X_1,Y) + \ldots + \lambda_k \mathrm{W}(X_k,Y)$







Équité algorithmique

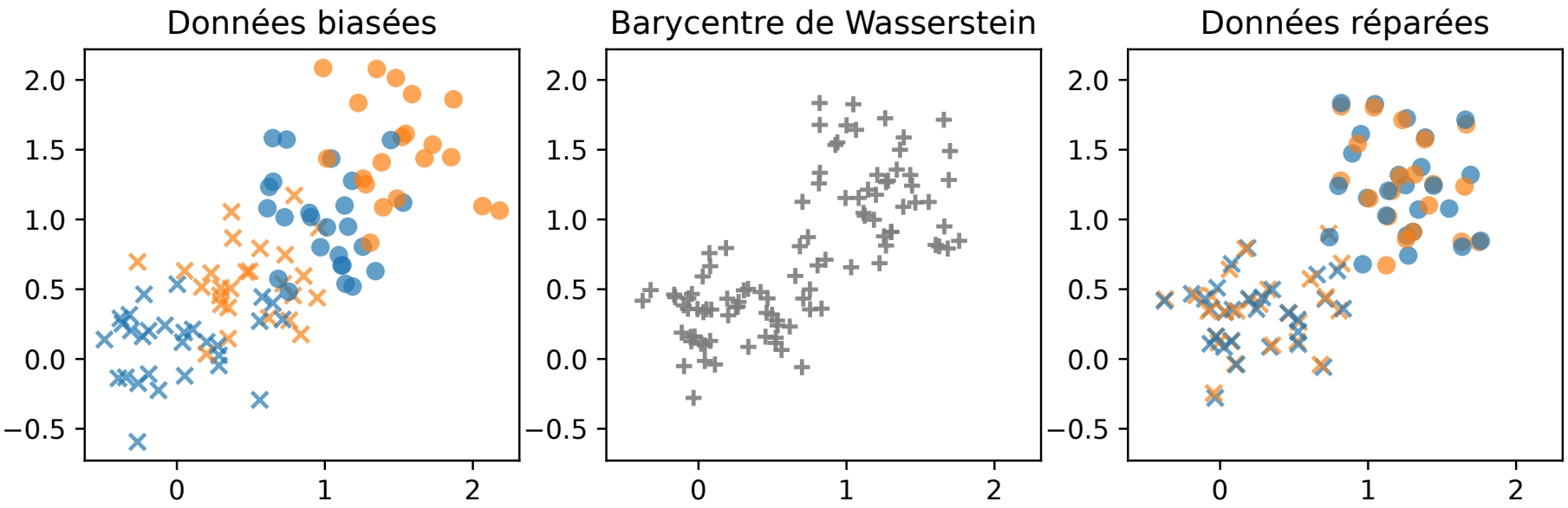
Effets des algorithmes dans la société

- Les IA sont apprises à partir de données qui peuvent être biaisées.
- Les algorithmes peuvent avoir des effets discriminatoires.

Exemple : COMPAS score en Floride, étudié par ProPublica

- Algorithme propriétaire prédisant le risques de récidive.
- Utilisé par les juges pour la mise en liberté avant les procés.
- Classé "à risque", sans récidive:
  - 23.5% des Blancs.
  - 44.5% des Afro-américains.
- Classé "sans risque", avec récidive:
  - 47.7% des Blancs.
  - 28.0% des Afro-américains.

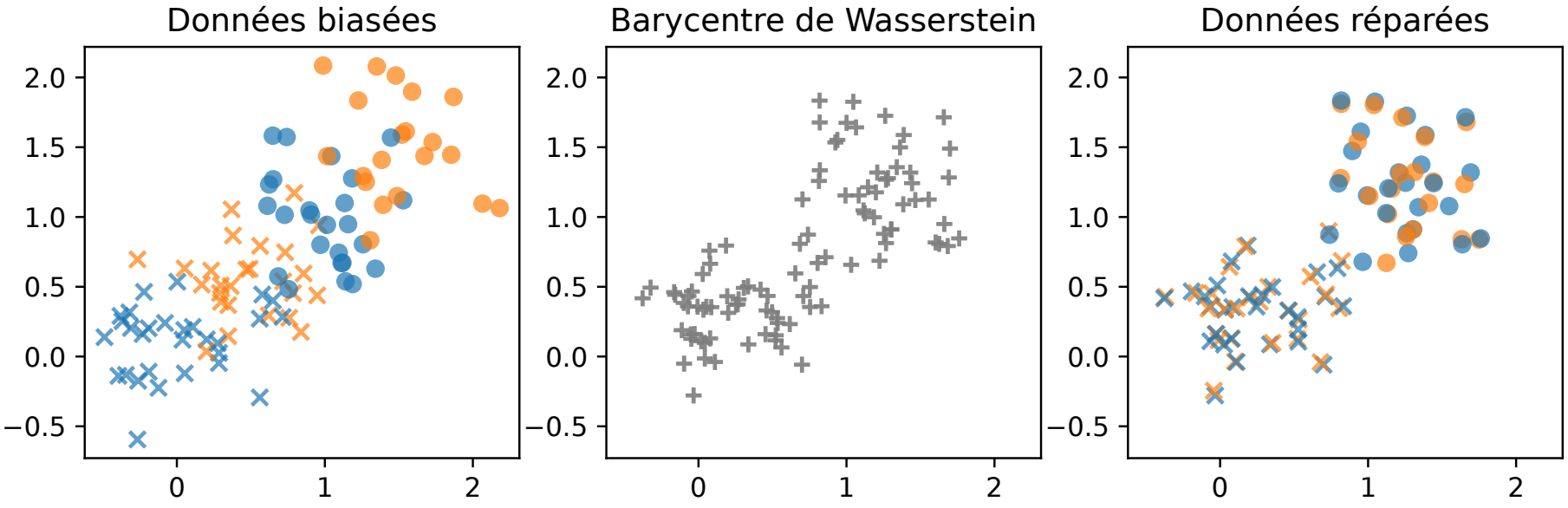
Transport Optimal pour l'équité algorithmique



Principe [Gordaliza et al., 2019]

- Séparer les données en deux distributions selon la valeur de la variable protégée.
- Calculer un barycentre de Wasserstein entre les deux distributions.
- Transporter les données de chaque distribution sur la barycentre pour obtenir un jeux de données "réparées" (adaptation de domaine vers le barycentre).
- Entraîner le classifieur sur les données réparées.
- Pour prédire sur de nouvelle donnée on applique le transport correspondant selon la variable protégée.

Transport Optimal pour l'équité algorithmique



Projet étudiant : équité avec transport optimal

B. Allouche

A. Ramanana Rahary

O. Rannen

Équité sur les données COMPAS

Apprentissage sur données COMPAS

- Précision du classifieur sur de nouvelles données : 65.6% (66% algo. propriétaire).
- Répartition des erreurs selon la population:

	Pred. Blancs		Pred. Afro-américains	
Récidive	Sans risque	A risque	Sans risque	A risque
Non	82.9%	17.1%	64.6%	35.4%
Oui	58.4%	41.6%	34.6%	65.4%

Classification équitable sur COMPAS

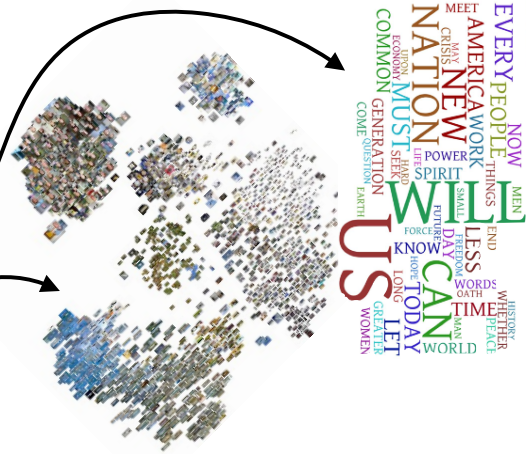
- Précision du classifieur équitable sur de nouvelles données : 61.8% (-3.8%).
- Répartition des erreurs selon la population:

	Pred. Blancs		Pred. Afro-américains	
Récidive	Sans risque	A risque	Sans risque	A risque
Non	86.4%	13.6%	88.2%	11.8%
Oui	68.7%	31.3%	67.1%	32.9%

- Biais compensé, petite perte de performance.
- En pratique on peut "réparer" en partie les données.

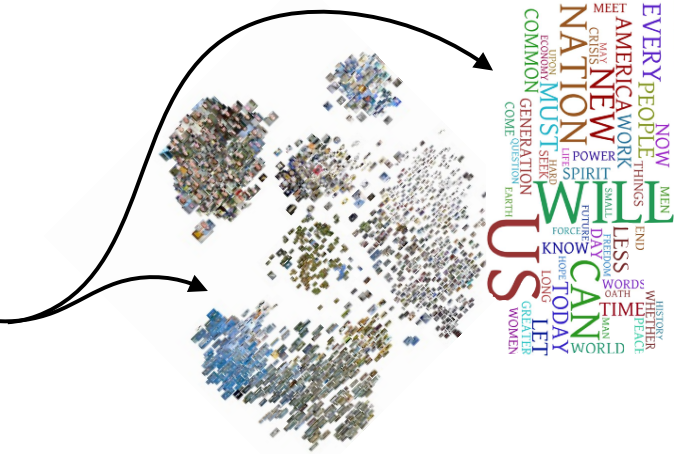
Conclusion

Le transport optimal : de la théorie à l'intelligence artificielle



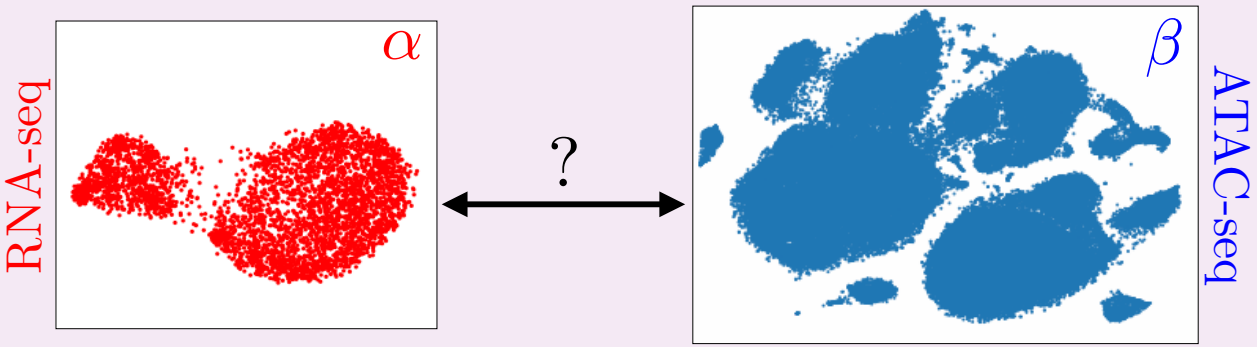
Conclusion

Le transport optimal : de la théorie à l'intelligence artificielle

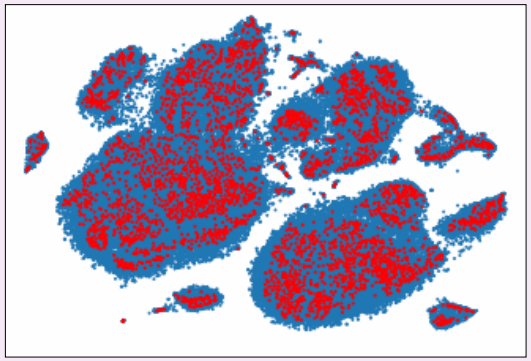


Le futur : transport optimal entre différents espaces !

[Othmane Sebbouh, 2021]



Multi-omics avec cellules uniques



Gromov-Wasserstein

References I

[Arjovsky et al., 2017] Arjovsky, M., Chintala, S., and Bottou, L. (2017).
Wasserstein generative adversarial networks.
In *Proceedings of the 34th International Conference on Machine Learning*, volume 70, pages 214–223, Sydney, Australia.

[Courty et al., 2016] Courty, N., Flamary, R., Tuia, D., and Rakotomamonjy, A. (2016).
Optimal transport for domain adaptation.
Pattern Analysis and Machine Intelligence, IEEE Transactions on.

[de Bézenac et al., 2021] de Bézenac, E., Ayed, I., and Gallinari, P. (2021).
Cyclegan through the lens of (dynamical) optimal transport.
In *Machine Learning and Knowledge Discovery in Databases. Research Track: European Conference, ECML PKDD 2021, Bilbao, Spain, September 13–17, 2021, Proceedings, Part II 21*, pages 132–147. Springer.

[Ferradans et al., 2014] Ferradans, S., Papadakis, N., Peyré, G., and Aujol, J.-F. (2014).
Regularized discrete optimal transport.
SIAM Journal on Imaging Sciences, 7(3).

[Goodfellow et al., 2014] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014).
Generative adversarial nets.
In *Advances in neural information processing systems*, pages 2672–2680.

References II

[Gordaliza et al., 2019] Gordaliza, P., Del Barrio, E., Fabrice, G., and Loubes, J.-M. (2019).
Obtaining fairness using optimal transport theory.
In *International Conference on Machine Learning*, pages 2357–2365. PMLR.

[Hoffman et al., 2018] Hoffman, J., Tzeng, E., Park, T., Zhu, J.-Y., Isola, P., Saenko, K., Efros, A., and Darrell, T. (2018).
Cycada: Cycle-consistent adversarial domain adaptation.
In *International conference on machine learning*, pages 1989–1998. Pmlr.

[Karras et al., 2019] Karras, T., Laine, S., and Aila, T. (2019).
A style-based generator architecture for generative adversarial networks.
In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4401–4410.

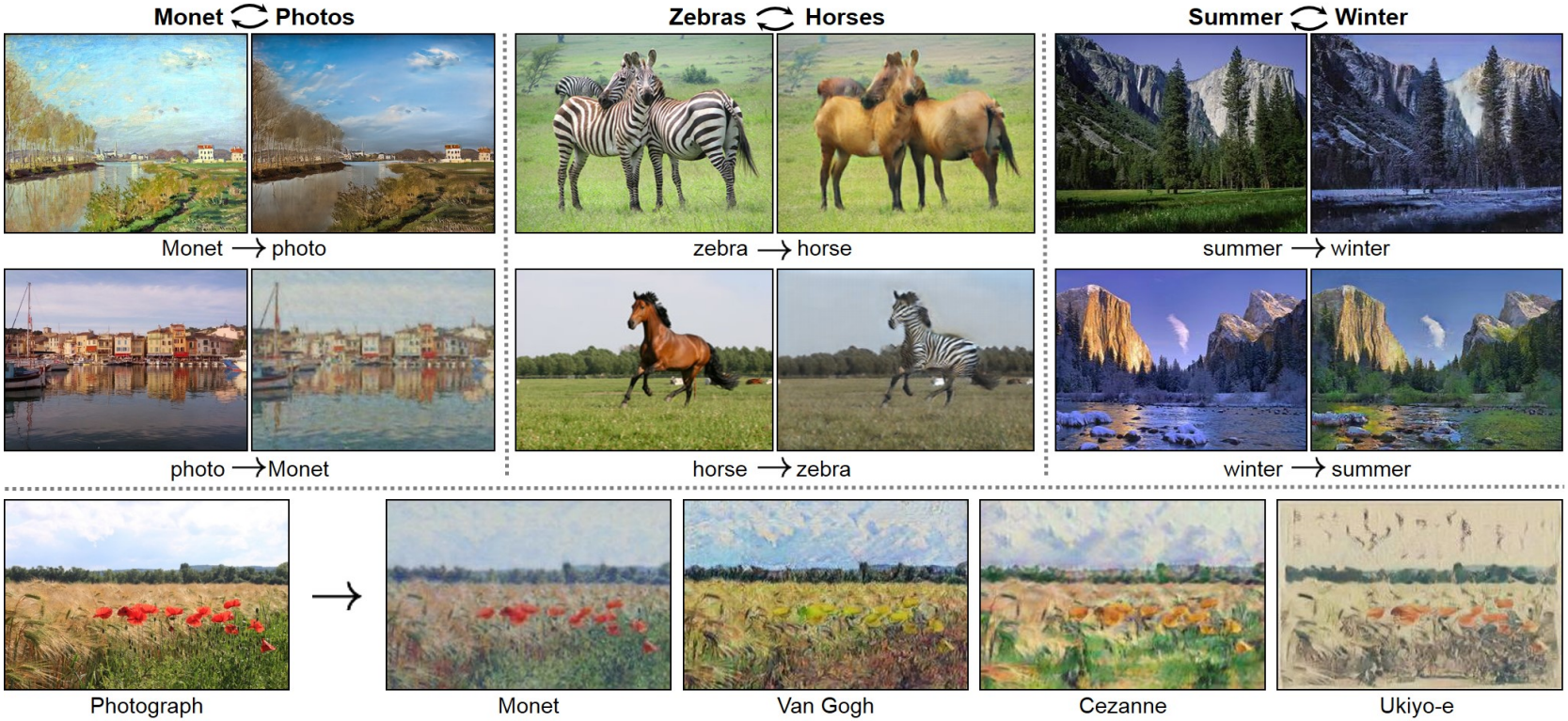
[Korotin et al., 2019] Korotin, A., Egiazarian, V., Asadulaev, A., Safin, A., and Burnaev, E. (2019).
Wasserstein-2 generative networks.
arXiv preprint arXiv:1909.13082.

[Mroueh, 2019] Mroueh, Y. (2019).
Wasserstein style transfer.
arXiv preprint arXiv:1905.12828.

References III

[Zhu et al., 2017] Zhu, J.-Y., Park, T., Isola, P., and Efros, A. A. (2017).
Unpaired image-to-image translation using cycle-consistent adversarial networks.
In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232.

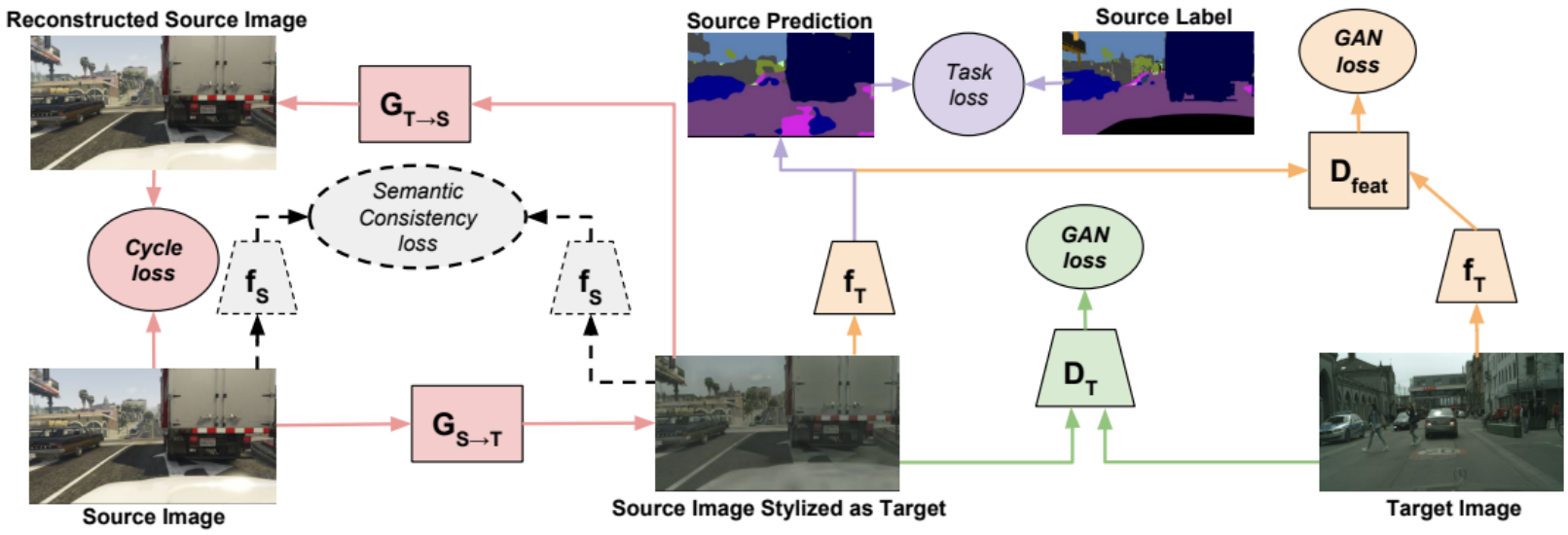
CycleGAN : GAN et adaptation de domaine



Principe [Zhu et al., 2017]

- Apprendre simultanément deux models génératifs source vers cible et cible vers source (pas de génération aléatoire).
- Peut être utilisé pour compenser des changements en adaptation de domaine [Hoffman et al., 2018].
- Les transformations CycleGAN entre données sont des approximations de transformations de transport optimal [de Bézenac et al., 2021].

CycleGAN : GAN et adaptation de domaine



Principe [Zhu et al., 2017]

- Apprendre simultanément deux models génératifs source vers cible et cible vers source (pas de génération aléatoire).
- Peut être utilisé pour compenser des changements en adaptation de domaine [Hoffman et al., 2018].
- Les transformations CycleGAN entre données sont des approximations de transformations de transport optimal [de Bézenac et al., 2021].