Deep Learning en Imágenes con GANs y Modelos de Difusión - Parte IV

Prof. Peter Montalvo

Agenda

CycleGANs 🔄

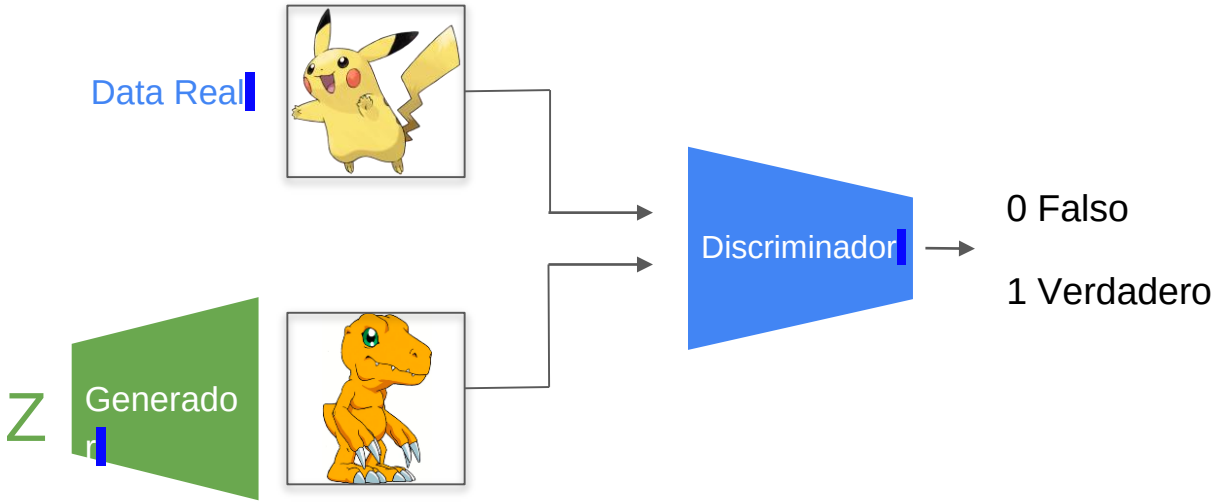
Modelos de Difusión □

Paired Data

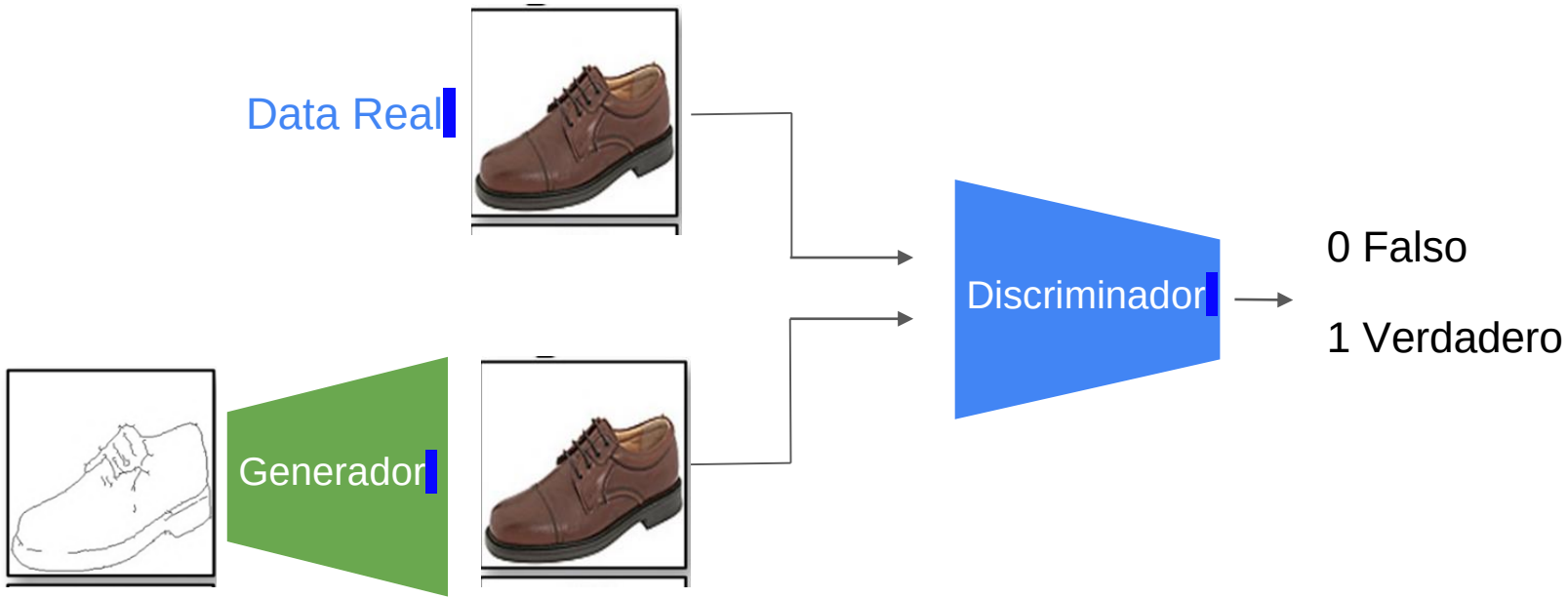
- Los datos emparejados son datos donde a cada x le corresponde un y.
- Esto en una GAN sería de la siguiente forma…

Paired Data


Paired
x_i
y_i


Data Real
Z
Generado
r
Discriminador
0 Falso
1 Verdadero

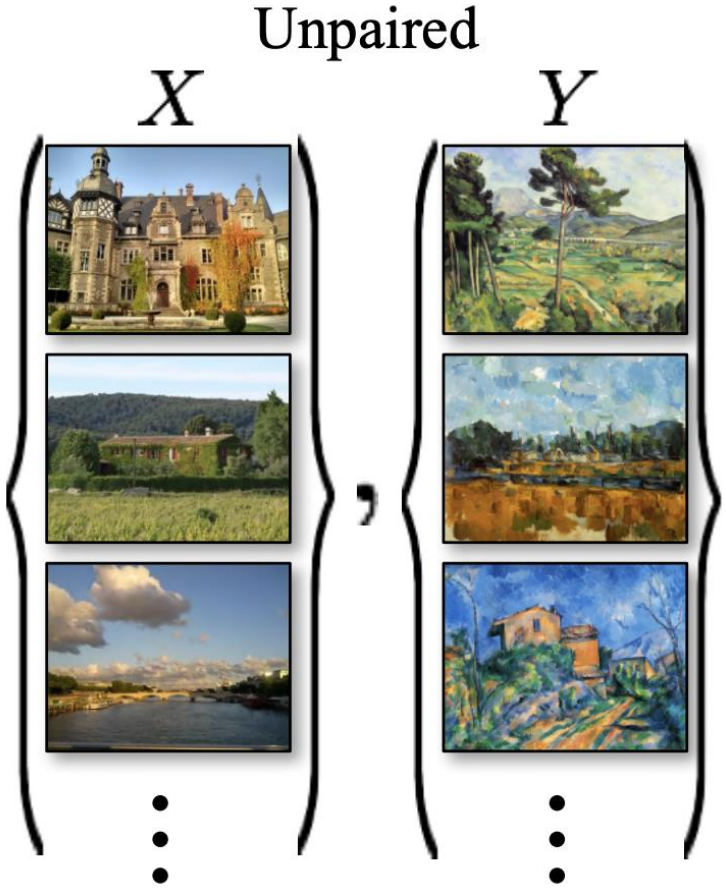
Paired Data



De forma que la GAN queda
de la siguiente forma…

Unpaired Data

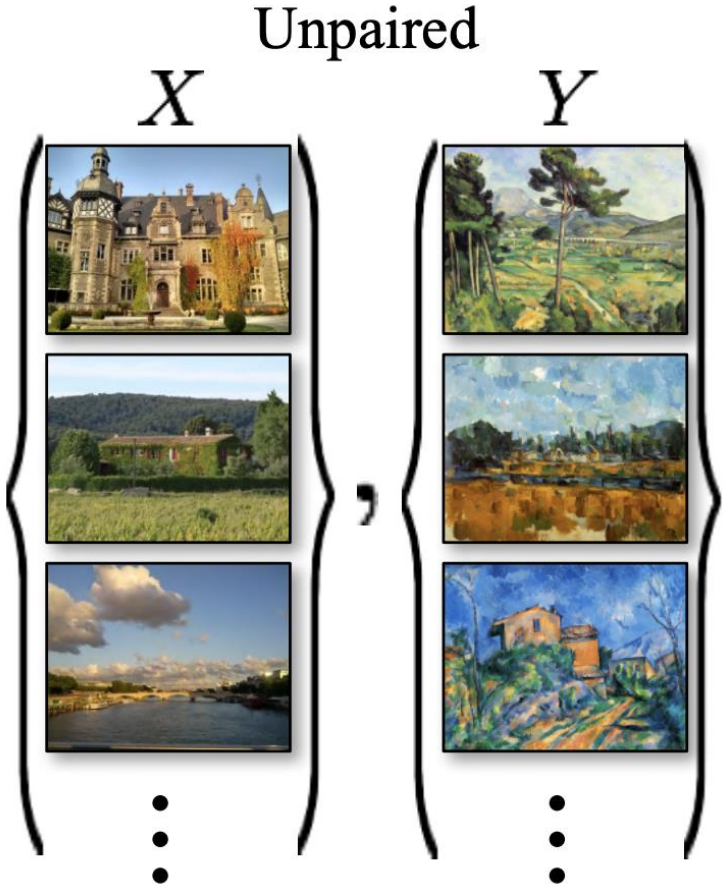
¿Qué pasaría si los datos no están emparejados?



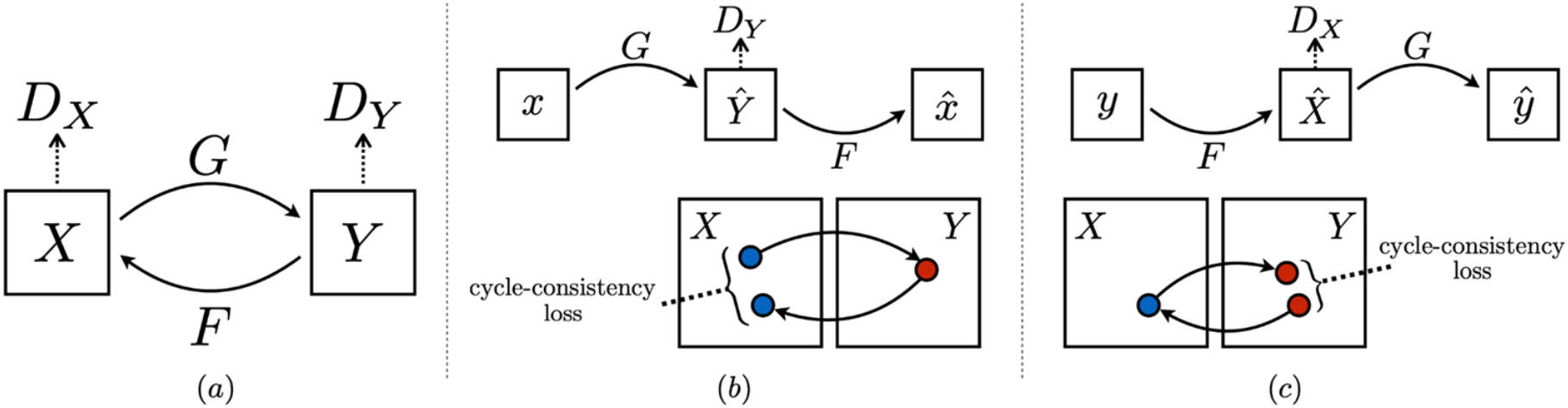
Unpaired Data

¿Qué pasaría si los datos no están emparejados?

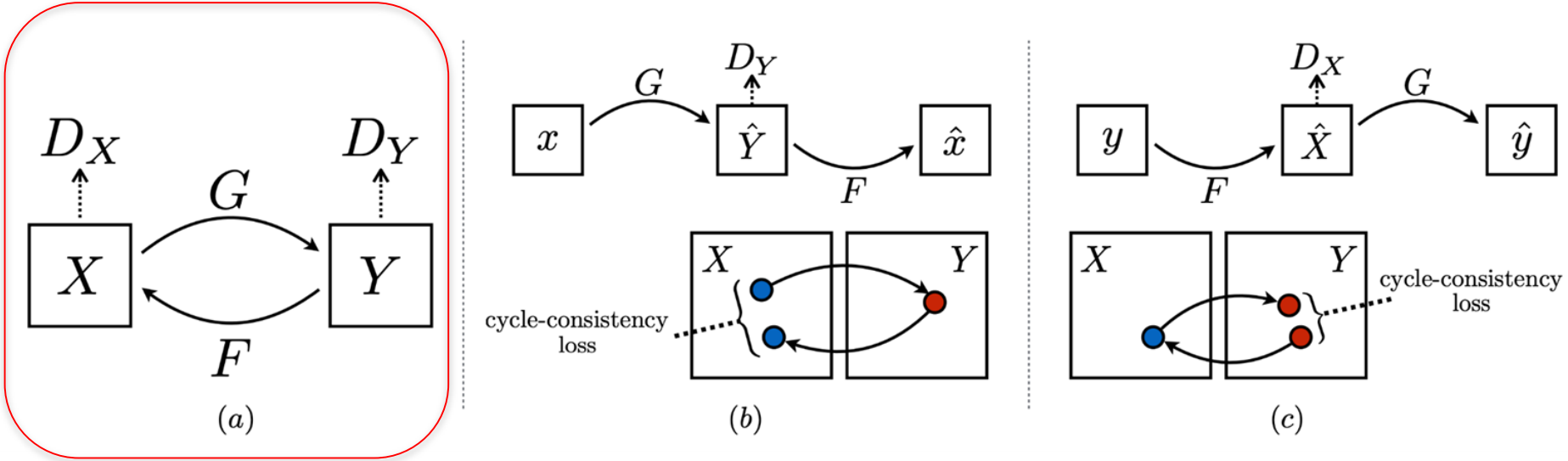
Acá es donde entran las CycleGANs 🔄



CycleGANs 🔄

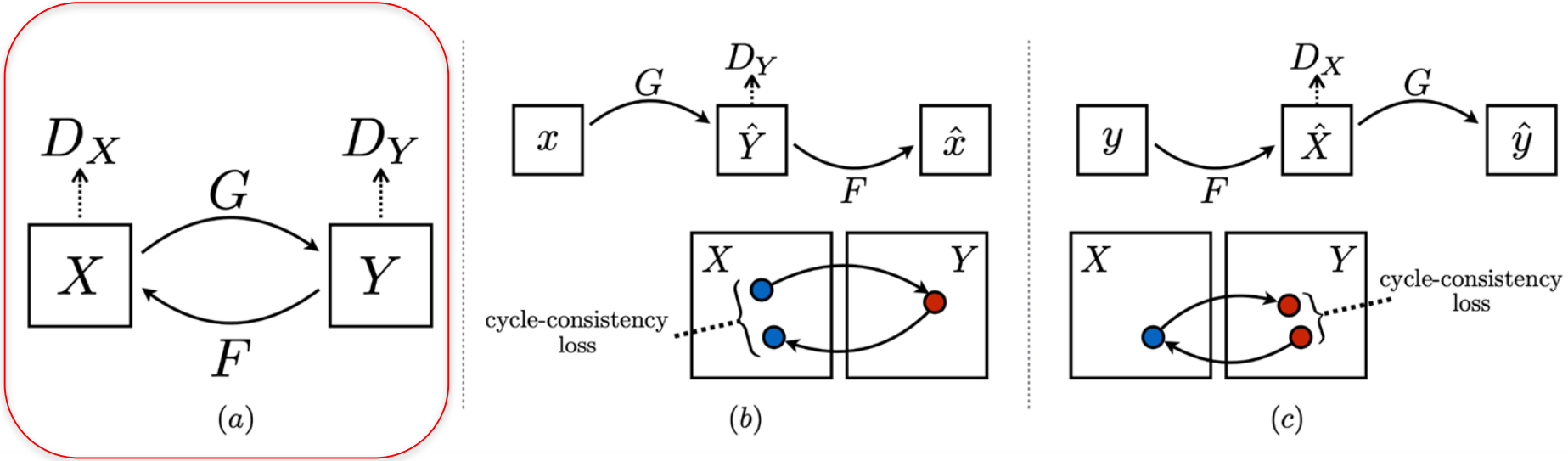

D_X
G
D_Y
X
Y
F
(a)
D_Y
G
x
$\hat{Y}$
$\hat{x}$
F
X
Y
cycle-consistency loss
(b)
D_X
G
y
$\hat{X}$
$\hat{y}$
F
X
Y
cycle-consistency loss
(c)

CycleGANs 🔄

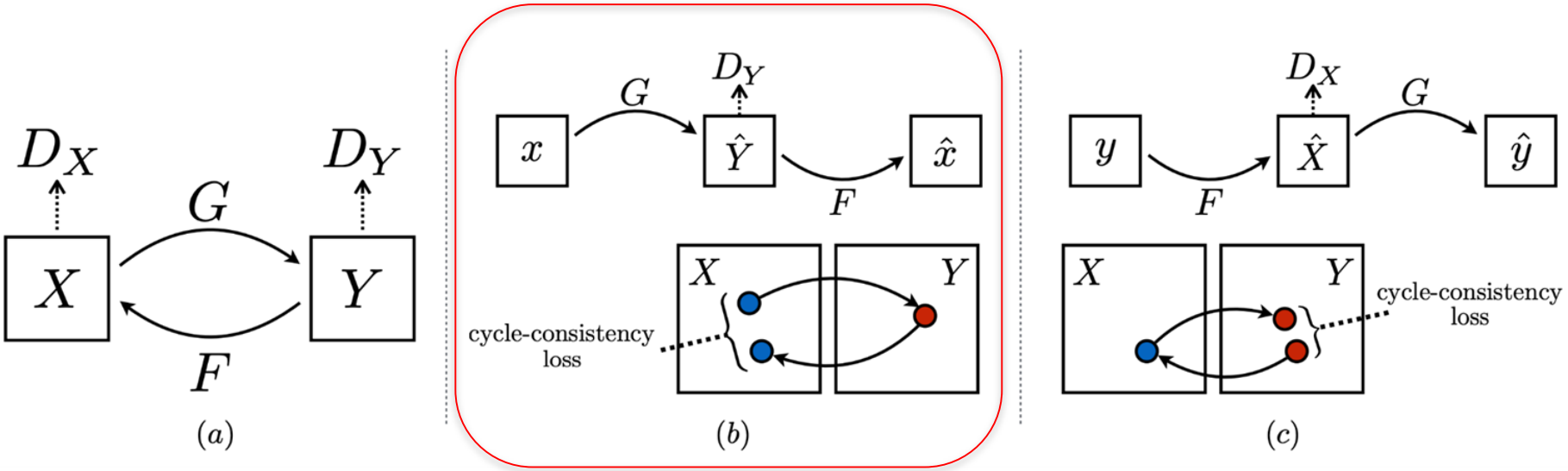

D_X
G
D_Y
X
Y
F
(a)
D_Y
G
x
$\hat{Y}$
$\hat{x}$
F
X
Y
cycle-consistency loss
(b)
D_X
G
y
$\hat{X}$
$\hat{y}$
F
X
Y
cycle-consistency loss
(c)

CycleGANs 🔄

G : X → Y y F : Y → X, y redes discriminadoras adversarias DY y DX

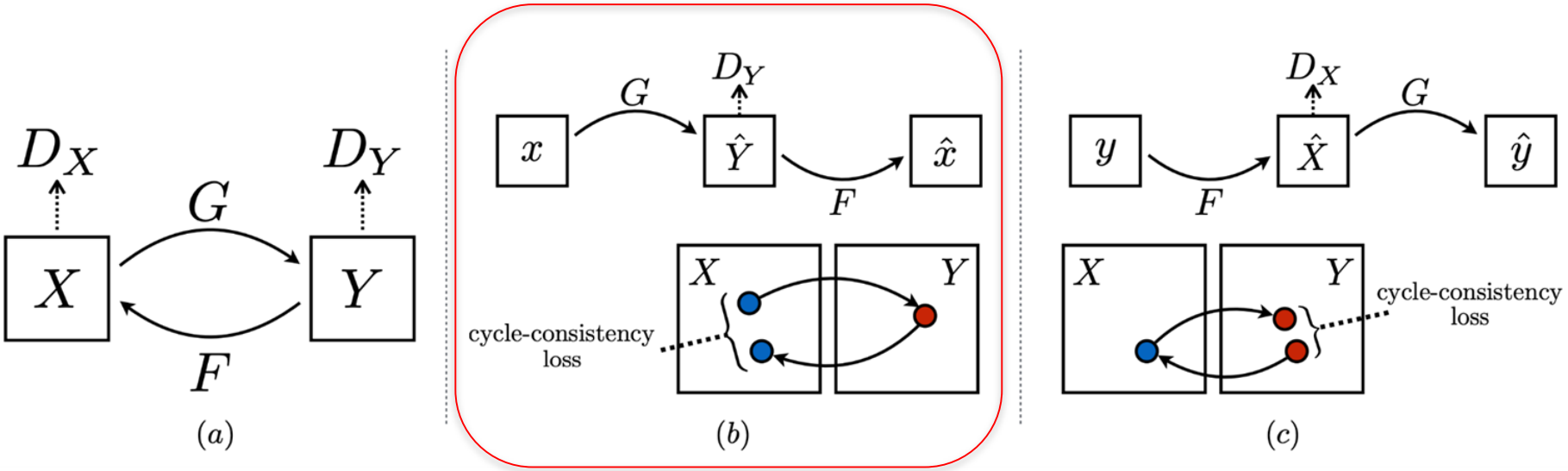


CycleGANs 🔄


D_X
G
D_Y
X
Y
F
(a)
x
G
D_Y
$\hat{Y}$
$\hat{x}$
F
X
Y
cycle-consistency loss
(b)
y
D_X
G
$\hat{X}$
$\hat{y}$
F
X
Y
cycle-consistency loss
(c)

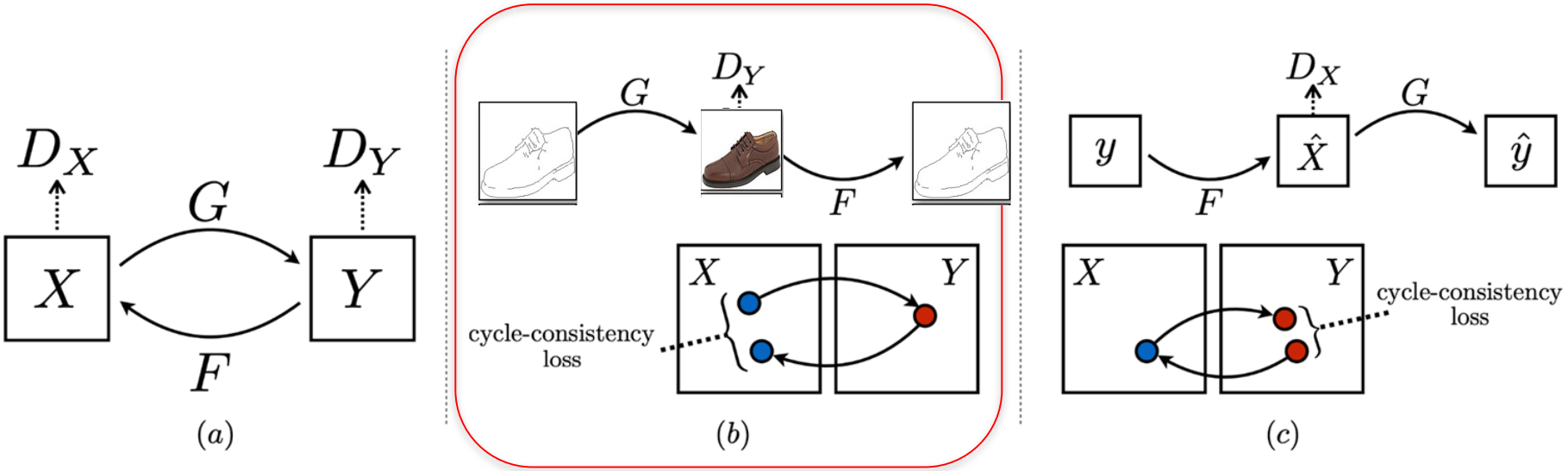
CycleGANs 🔄

G va de X a Y, mientras que F va de Y a X.



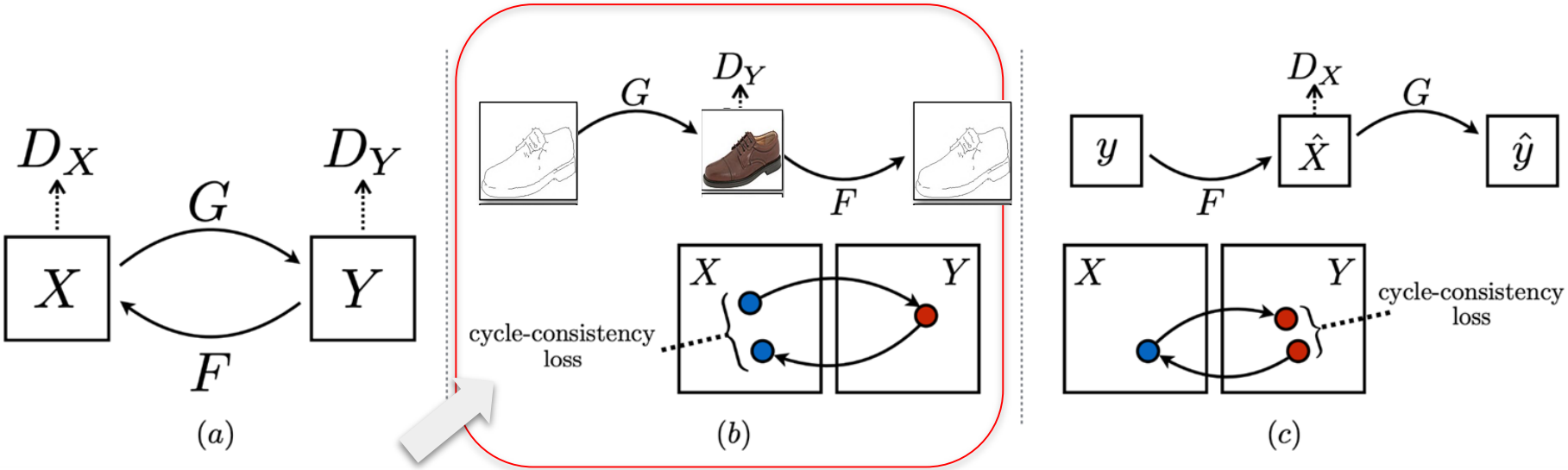
CycleGANs 🔄

Gráficamente…



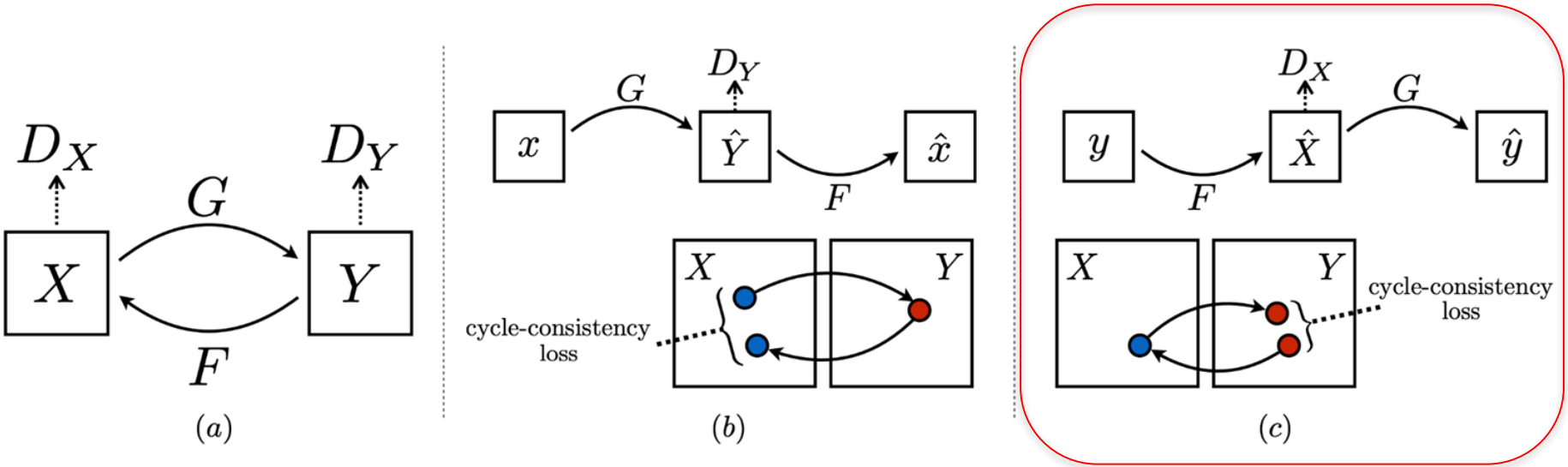
CycleGANs 🔄

Nótese que hay una función de pérdida llamada “Cycle Consistency Loss”



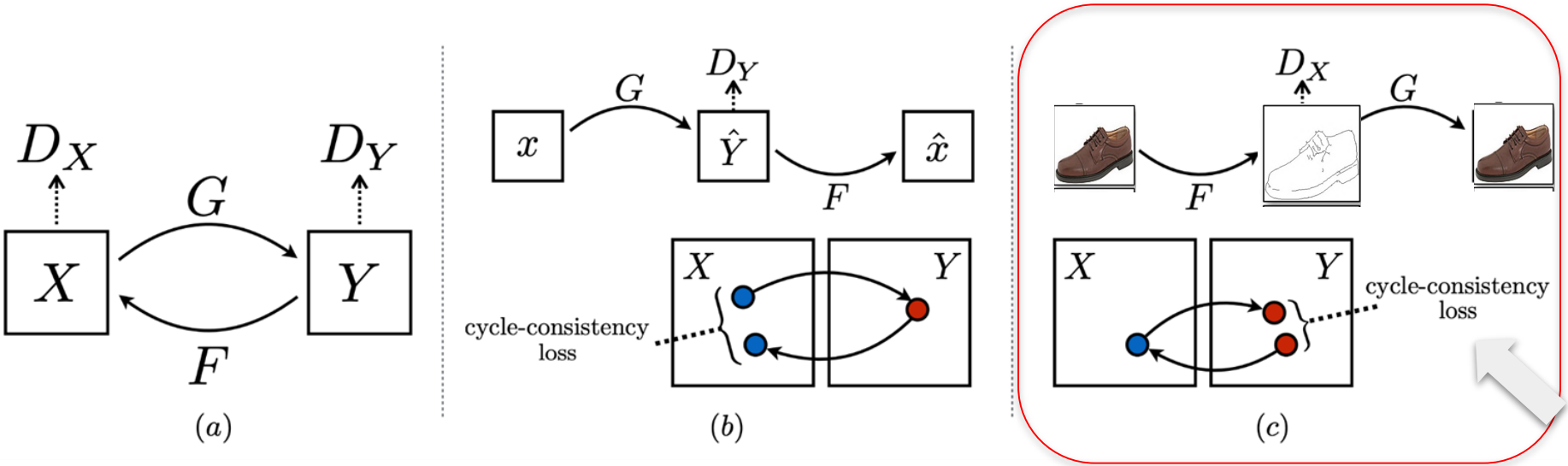
CycleGANs 🔄

Y también de viceversa



CycleGANs 🔄

Y también de viceversa

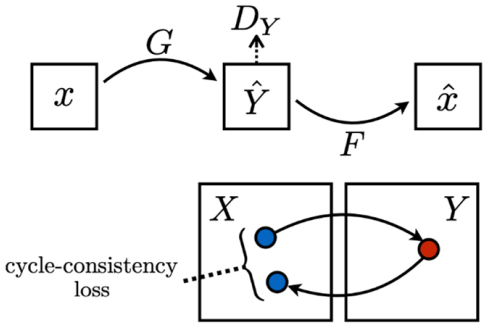


CycleGANs 🔄 - Función de pérdida

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1].$$

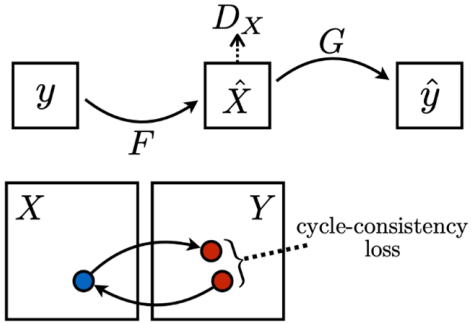
CycleGANs 🔄 - Función de pérdida

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1].$$



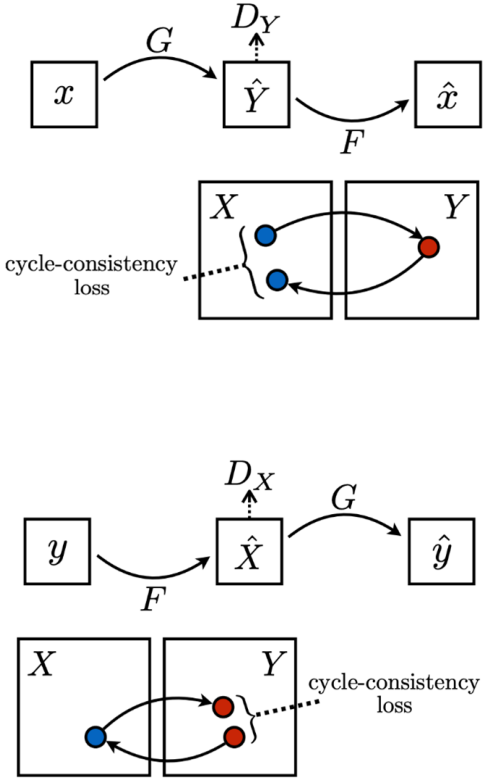
CycleGANs 🔄 - Función de pérdida

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1].$$



CycleGANs 🔄 - Función de pérdida

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1].$$



CycleGANs 🔄 - Función de pérdida

De forma que la función de pérdida general queda de la siguiente forma…

$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) =& \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ &+ \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ &+ \lambda \mathcal{L}_{\text{cyc}}(G, F),\end{aligned}$$

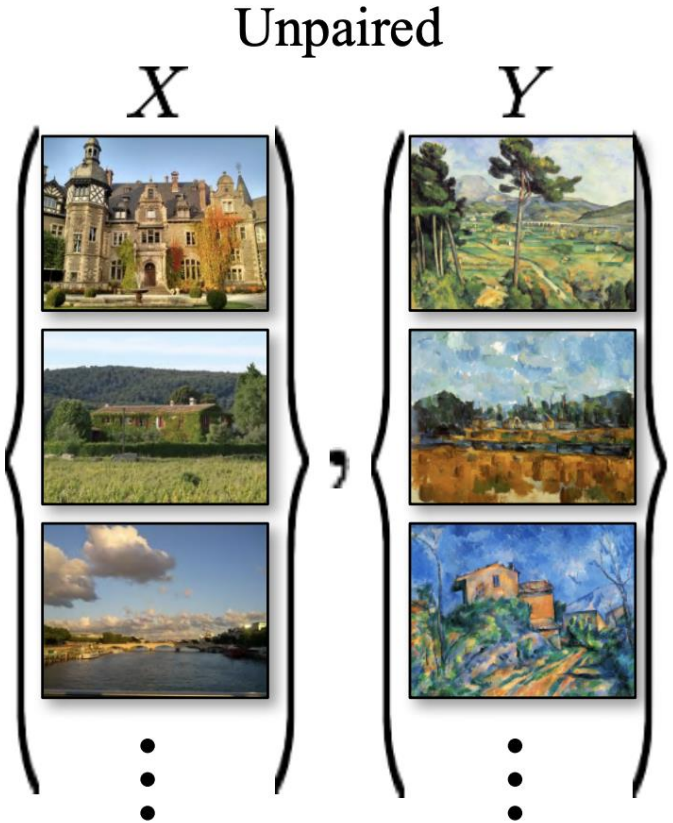
CycleGANs 🔄 - Función de pérdida

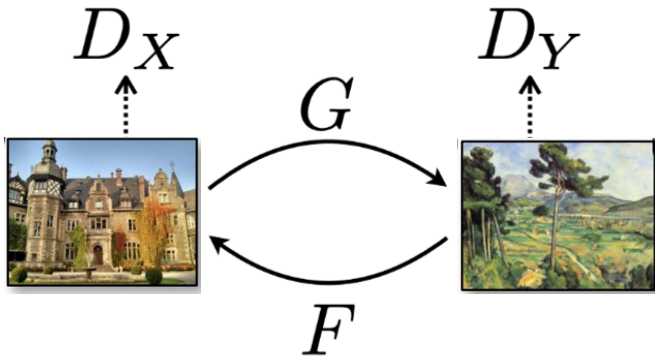
De forma que la función de pérdida general queda de la siguiente forma…

$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) =& \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ &+ \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ &+ \lambda \mathcal{L}_{\text{cyc}}(G, F),\end{aligned}$$

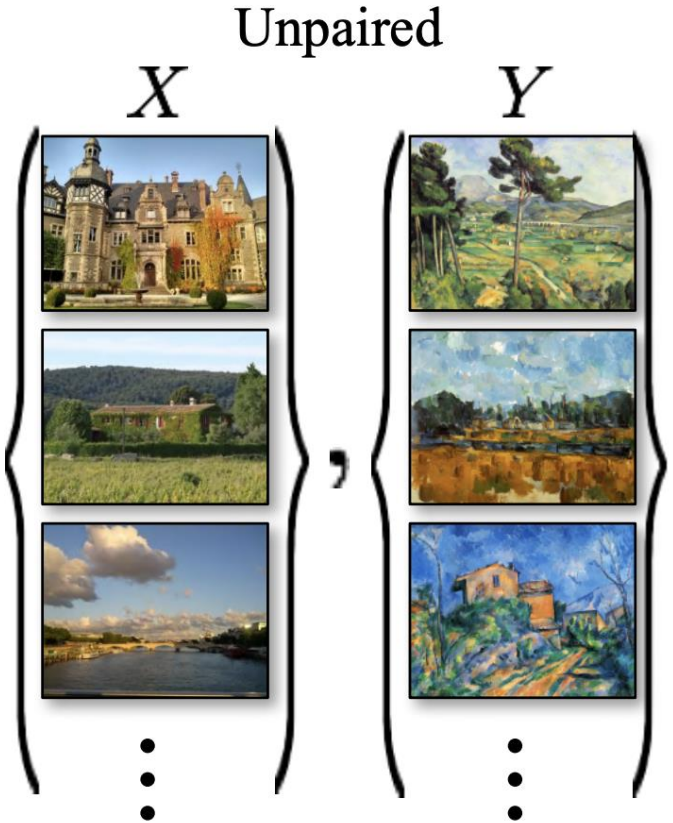
$$\begin{aligned}\mathcal{L}_{\text{cyc}}(G, F) = & \mathbb{E}_{x \sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] \\ &+ \mathbb{E}_{y \sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1].\end{aligned}$$

CycleGANs 🔄 - Datos no emparejada

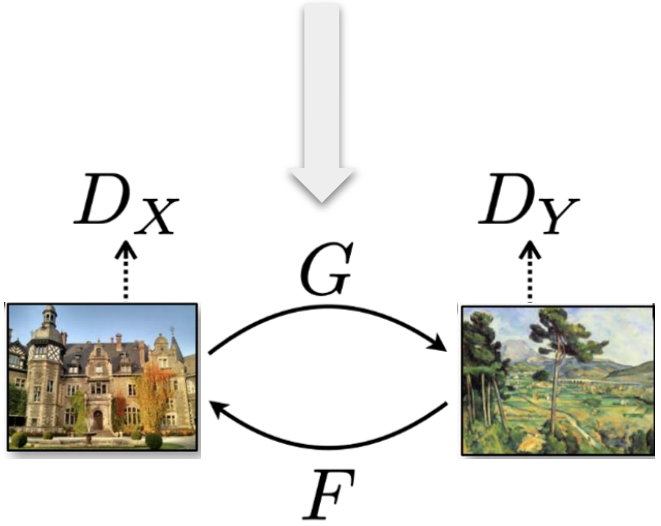




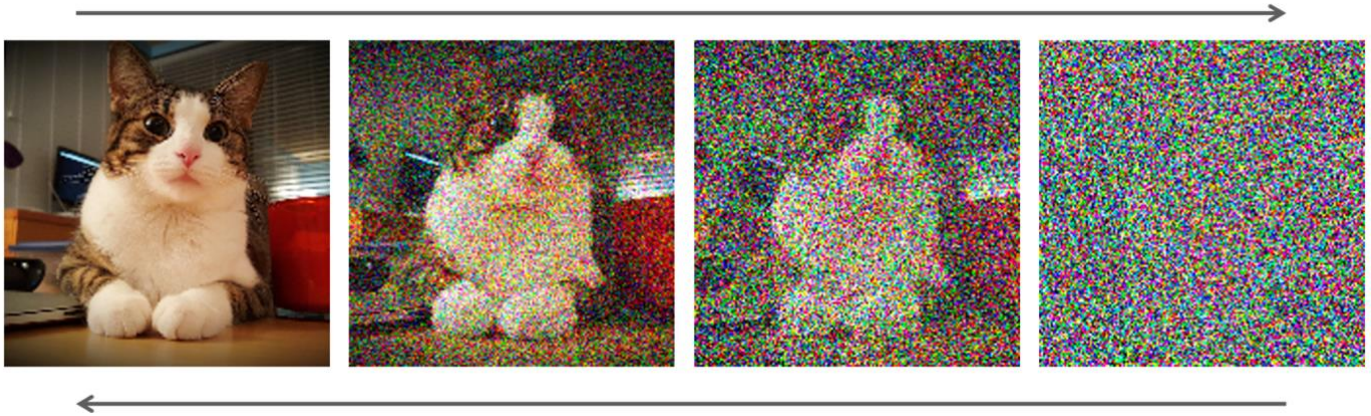
CycleGANs 🔄 - Datos no emparejada



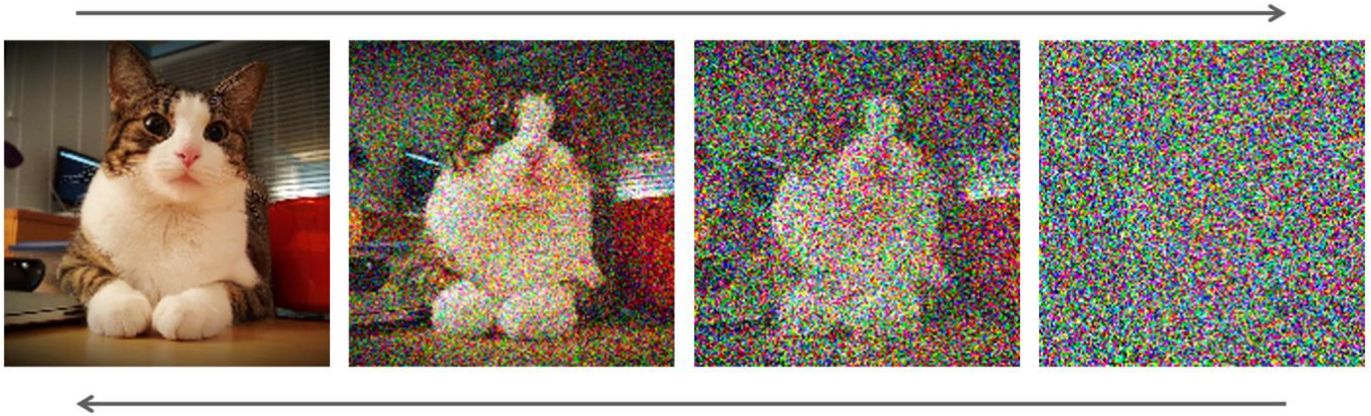
Esto permite una traducción de imágenes con diversos estilos



Modelos de Difusión □



Modelos de Difusión □



demo en:
https://imagen.research.google/

Modelos de Difusión

Distancia de inicio de Fréchet

Es una métrica utilizada para determinar la calidad de las imágenes creadas por un modelo generativo, como una red adversarial generativa (GAN).

A diferencia de la puntuación de inicio anterior (IS), que evalúa solo la distribución de las imágenes generadas, la DIF compara la distribución de las imágenes generadas con la distribución de un conjunto de imágenes reales («verdad fundamental»).

Más info:
https://es.wikipedia.org/wiki/Distancia_de_inicio_de_Fr%C3%A9chet

Modelos de Difusión □

Figure 6: Samples from BigGAN-deep with truncation 1.0 (FID 6.95, left) vs samples from our diffusion model with guidance (FID 4.59, middle) and samples from the training set (right).

Más info:
https://es.wikipedia.org/wiki/Distancia_de_inicio_de_Fr%C3%A9chet

Modelos de Difusión □

FID 6.95

FID 4.59

Más info:
https://es.wikipedia.org/wiki/Distancia_de_inicio_de_Fr%C3%A9chet

Modelos de Difusión □

FID 6.95

GAN

FID 4.59

DM

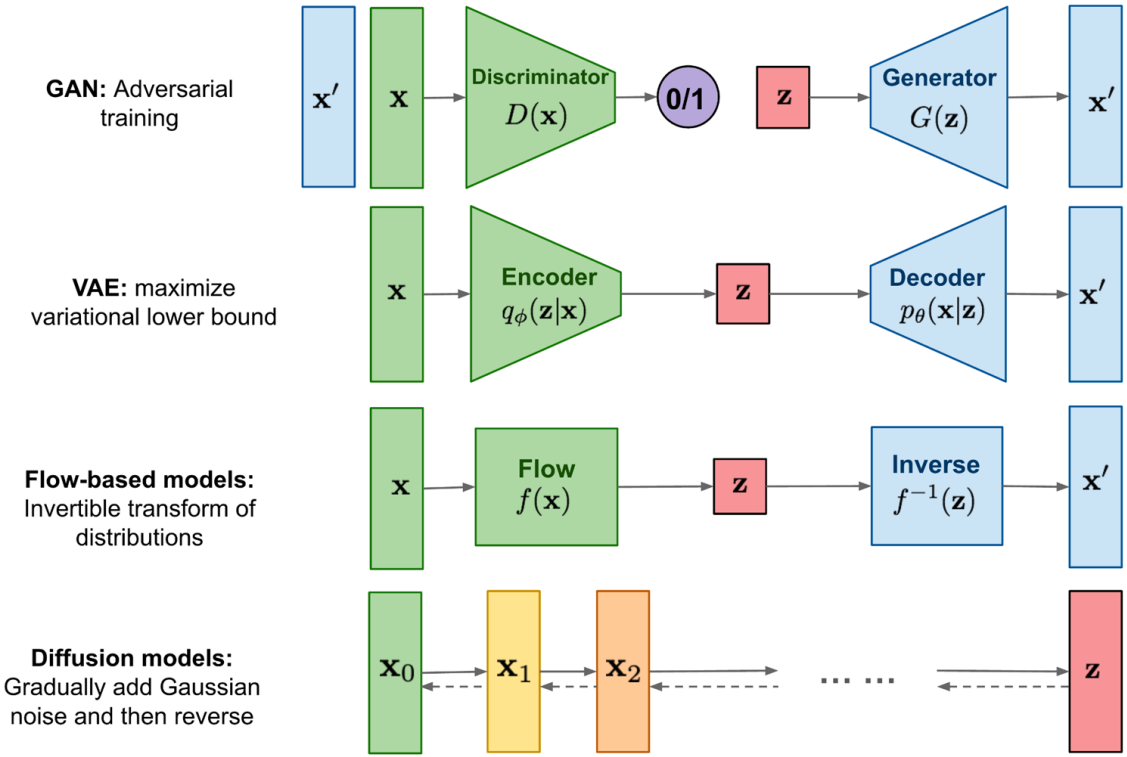
Más info:
https://es.wikipedia.org/wiki/Distancia_de_inicio_de_Fr%C3%A9chet

Modelos de Difusión

- Los modelos GAN son conocidos por su entrenamiento potencialmente inestable y su menor diversidad generacional debido a su naturaleza de entrenamiento adversario.

- VAE se basa en una pérdida sustituta. Los modelos de flujo deben utilizar arquitecturas especializadas para construir transformaciones reversibles.

Modelos de Difusión

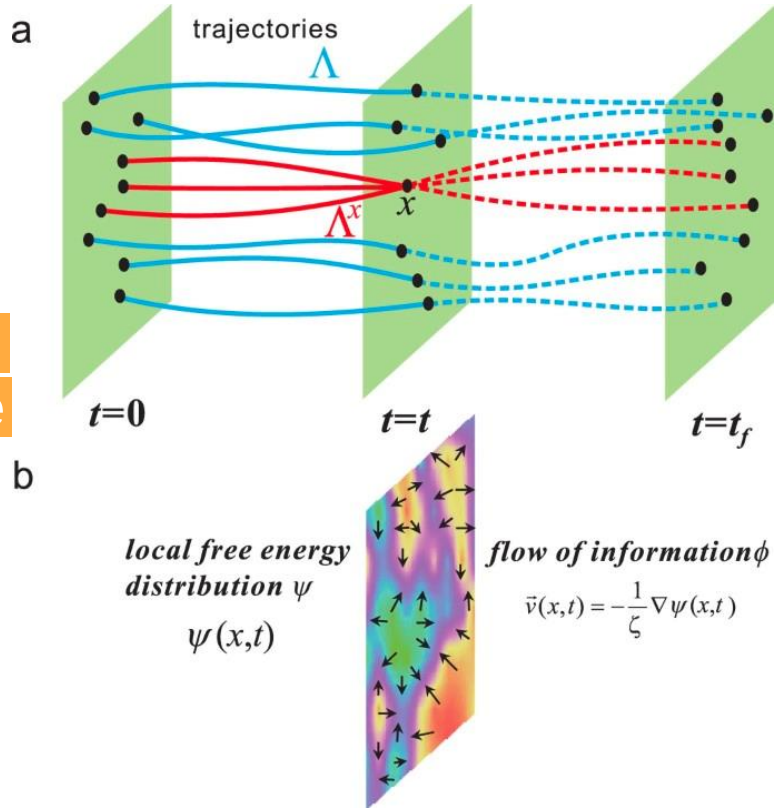


Modelos de Difusión

- Los modelos de difusión están inspirados en la termodinámica de no equilibrio.

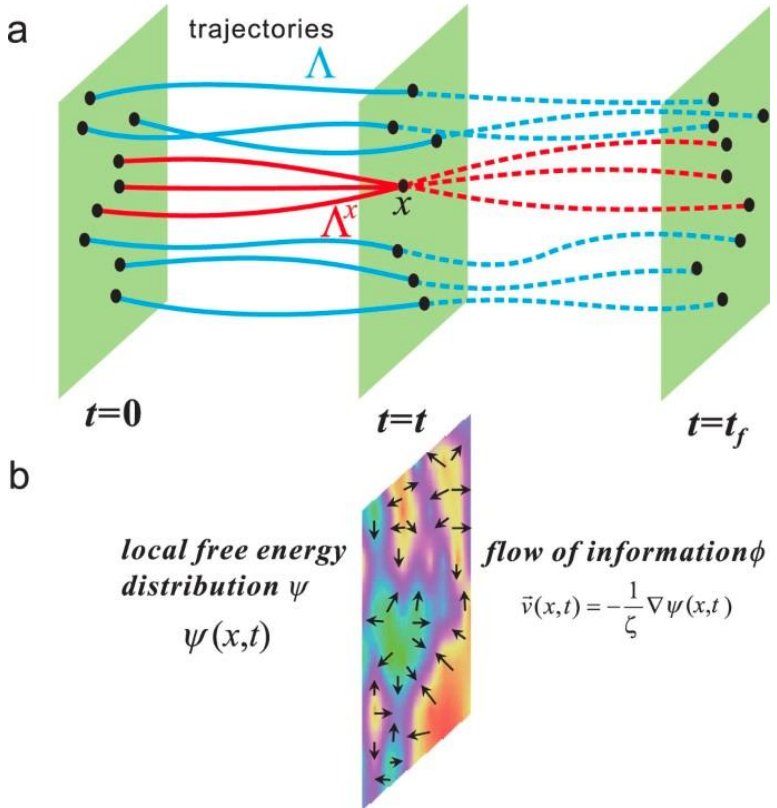
Modelos de Difusión

- Tienen una cadena de Markov de pasos de difusión para agregar lentamente ruido aleatorio a los datos y luego aprenden a revertir el proceso de difusión para construir muestras de datos deseadas a partir del ruido.



Modelos de Difusión

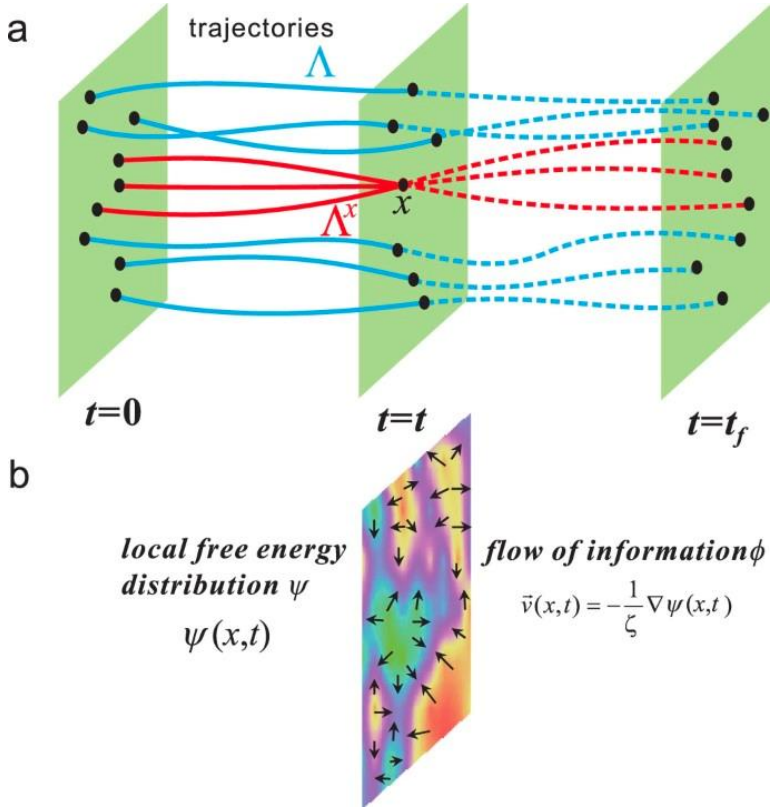
- A diferencia de los VAE o los modelos de flujo, los modelos de difusión se aprenden con un procedimiento fijo y su espacio latente tiene una alta dimensionalidad (igual que los datos originales).



Modelos de Difusión

Predecesores:

- Modelos probabilísticos de difusión (Sohl-Dickstein et al., 2015)
- Redes de puntuación condicionadas por ruido (NCSN ; Yang y Ermon, 2019)
- Modelos probabilísticos de difusión con eliminación de ruido (DDPM ; Ho et al.2020).



Modelos de Difusión - Cómo funciona

Proceso de difusión directa

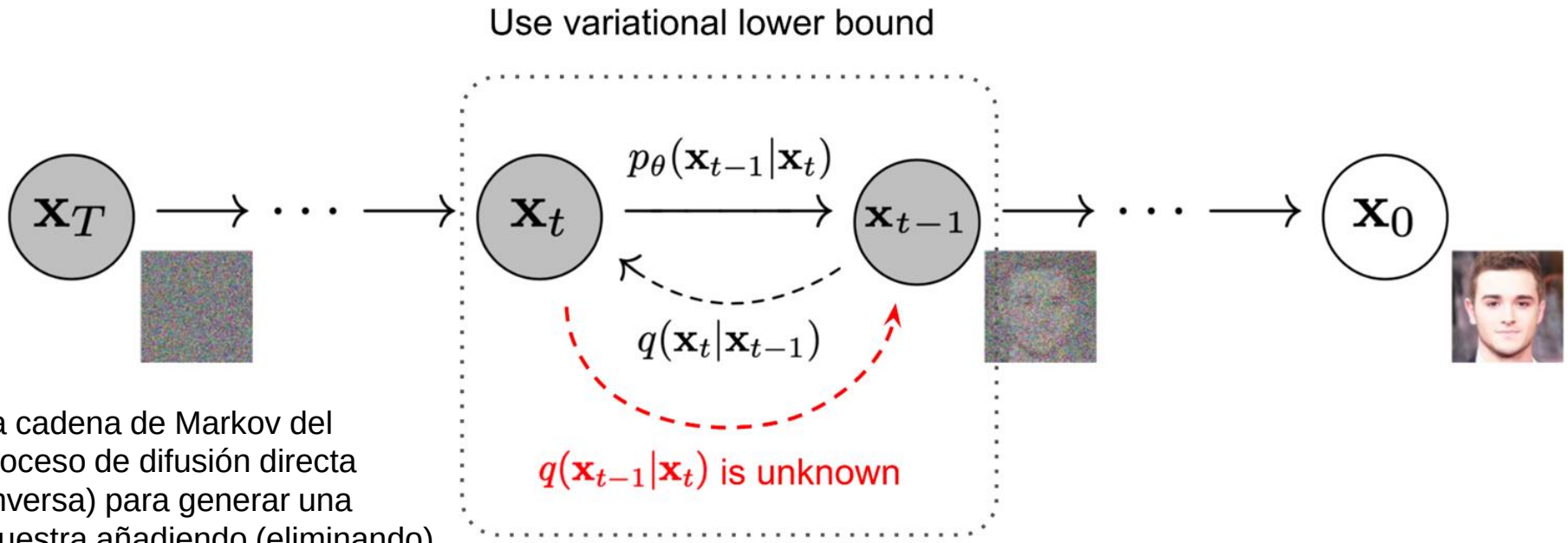
Dado un punto de datos muestreado de una distribución de datos real$\mathbf{x}_0 \sim q(\mathbf{x})$, definamos un *proceso de difusión directa* en el que agregamos una pequeña cantidad de ruido gaussiano a la muestra en$T$pasos, produciendo una secuencia de muestras ruidosas$\mathbf{x}_1, \dots, \mathbf{x}_T$. Los tamaños de los pasos están controlados por un programa de variación.$\{\beta_t \in (0,1)\}_{t=1}^T$.

Modelos de Difusión - Cómo funciona

Consideremos la siguiente forma para q:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}) \quad q(\mathbf{x}_{1:T}|\mathbf{x}_0) = \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1})$$

Modelos de Difusión - Cómo funciona



La cadena de Markov del proceso de difusión directa (inversa) para generar una muestra añadiendo (eliminando) ruido lentamente.

Modelos de Difusión - Cómo funciona

Una buena propiedad del proceso anterior es que podemos muestrear$\mathbf{x}_t$en cualquier paso de tiempo arbitrario$t$en forma cerrada usando el truco de reparametrización . Dejar$\alpha_t = 1 - \beta_t$y $\bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i$:

$$
\begin{aligned}
\mathbf{x}_t &= \sqrt{\alpha_t}\mathbf{x}_{t-1} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_{t-1} && ;\text{where } \boldsymbol{\epsilon}_{t-1}, \boldsymbol{\epsilon}_{t-2}, \cdots \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \\
&= \sqrt{\alpha_t\alpha_{t-1}}\mathbf{x}_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\bar{\boldsymbol{\epsilon}}_{t-2} && ;\text{where } \bar{\boldsymbol{\epsilon}}_{t-2} \text{ merges two Gaussians (*).} \\
&= \dots \\
&= \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon} \\
q(\mathbf{x}_t|\mathbf{x}_0) &= \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})
\end{aligned}
$$

(*) Recuerde que cuando fusionamos dos gaussianos con diferente varianza,$\mathcal{N}(\mathbf{0}, \sigma_1^2\mathbf{I})$y $\mathcal{N}(\mathbf{0}, \sigma_2^2\mathbf{I})$, la nueva distribución es$\mathcal{N}(\mathbf{0}, (\sigma_1^2 + \sigma_2^2)\mathbf{I})$. Aquí la desviación estándar combinada es $\sqrt{(1-\alpha_t) + \alpha_t(1-\alpha_{t-1})} = \sqrt{1-\alpha_t\alpha_{t-1}}$.

Modelos de Difusión - Cómo funciona

Conexión con la dinámica de Langevin del gradiente estocástico.

La dinámica de Langevin es un concepto de la física, desarrollado para modelar estadísticamente sistemas moleculares. Combinada con el descenso del gradiente estocástico, *la dinámica de Langevin del gradiente estocástico* (Welling & Teh 2011) puede producir muestras a partir de una densidad de probabilidad.$p(\mathbf{x})$usando solo los gradientes$\nabla_{\mathbf{x}} \log p(\mathbf{x})$en una cadena de actualizaciones de Markov:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \frac{\delta}{2}\nabla_{\mathbf{x}} \log p(\mathbf{x}_{t-1}) + \sqrt{\delta}\boldsymbol{\epsilon}_t, \quad \text{where } \boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

dóndeδes el tamaño del paso. Cuando$T \to \infty, \epsilon \to 0,\mathbf{x}_T$es igual a la verdadera densidad de probabilidad$p(\mathbf{x})$.

En comparación con el SGD estándar, la dinámica de Langevin con gradiente estocástico inyecta ruido gaussiano en las actualizaciones de parámetros para evitar colapsos en los mínimos locales.

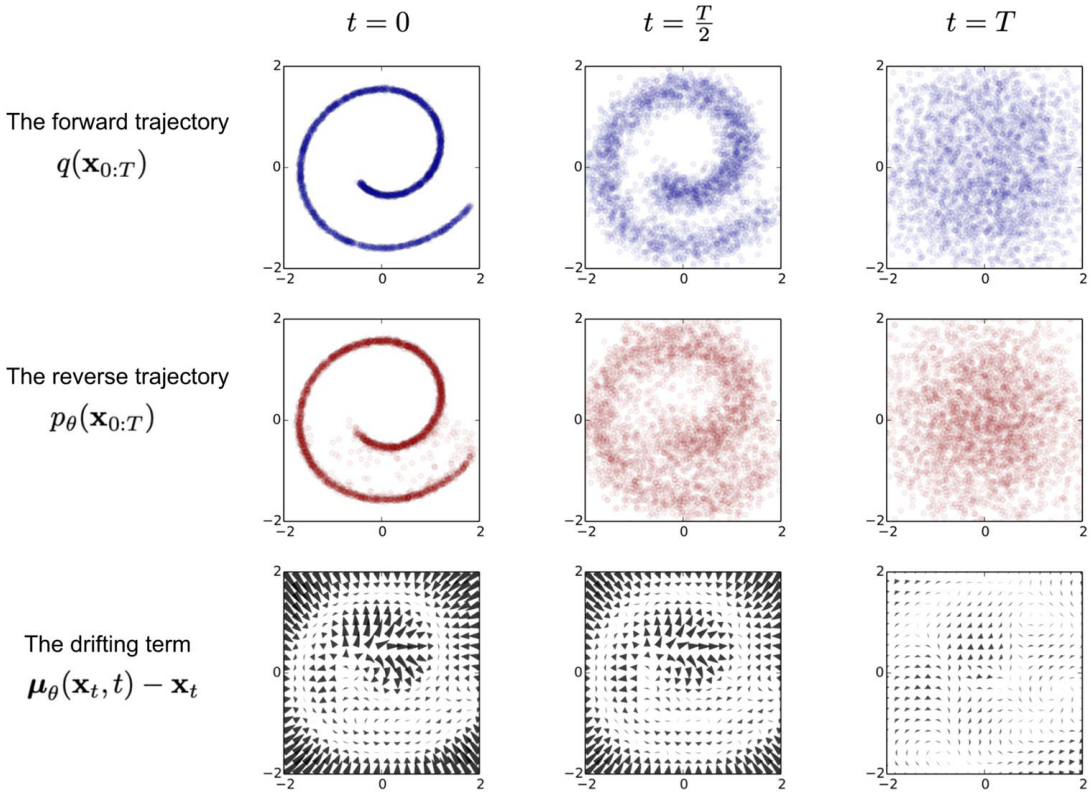
Modelos de Difusión - Cómo funciona

Proceso de difusión inversa

Si podemos revertir el proceso anterior y tomar muestras de$q(\mathbf{x}_{t-1}|\mathbf{x}_t)$, podremos recrear la muestra real a partir de una entrada de ruido gaussiano,$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Tenga en cuenta que si$\beta_t$ es lo suficientemente pequeño,$q(\mathbf{x}_{t-1}|\mathbf{x}_t)$también será gaussiano. Desafortunadamente, no podemos estimar fácilmente$q(\mathbf{x}_{t-1}|\mathbf{x}_t)$porque necesita utilizar todo el conjunto de datos y, por lo tanto, necesitamos aprender un modelo$p_\theta$aproximar estas probabilidades condicionales para ejecutar el *proceso de difusión inversa* .

$$p_\theta(\mathbf{x}_{0:T}) = p(\mathbf{x}_T) \prod_{t=1}^{T} p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) \quad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, t), \boldsymbol{\Sigma}_\theta(\mathbf{x}_t, t))$$

Modelos de Difusión - Cómo funciona



Fuentes

- https://arxiv.org/abs/1703.10593 - CycleGANs
- https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-1/
- https://developer.nvidia.com/blog/improving-diffusion-models-as-an-alternative-to-gans-part-2/
- https://arxiv.org/abs/2105.05233 - Diffusion Models
- https://arxiv.org/abs/2006.11239 - Denoising Probabilistic Models
- https://lilianweng.github.io/posts/2021-07-11-diffusion-models/ - What are dm