UTEC
UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

> Reinventa el mundo <

Sesión 2.2

Training and Representations

Tricks, Generalización

TRANSFORMATEC

UTEC
UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

1.

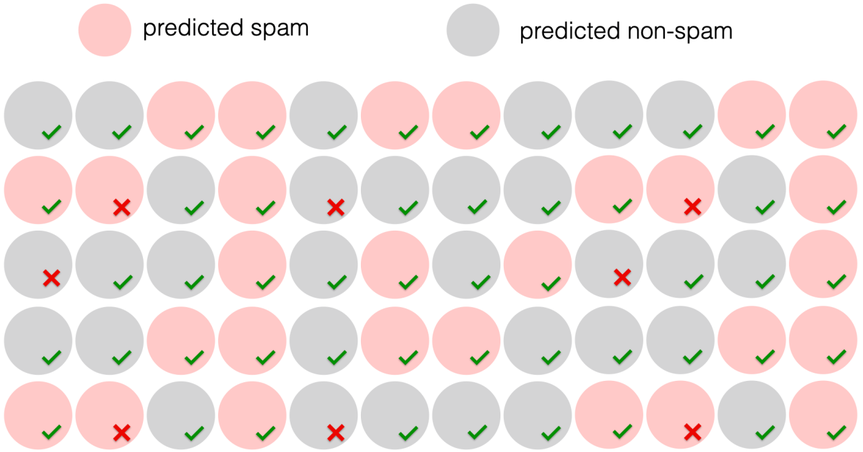
Metrics

TRANSFORMATEC

Métricas

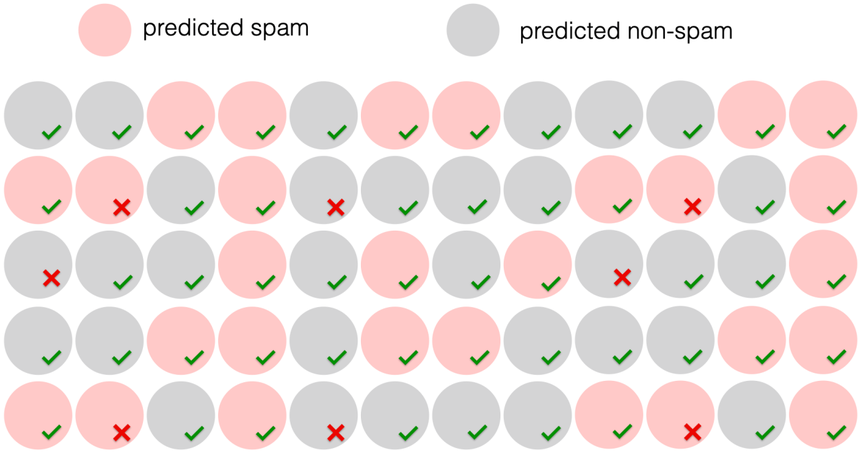
		Actual	
		Positive	Negative
Predicted	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

Accuracy



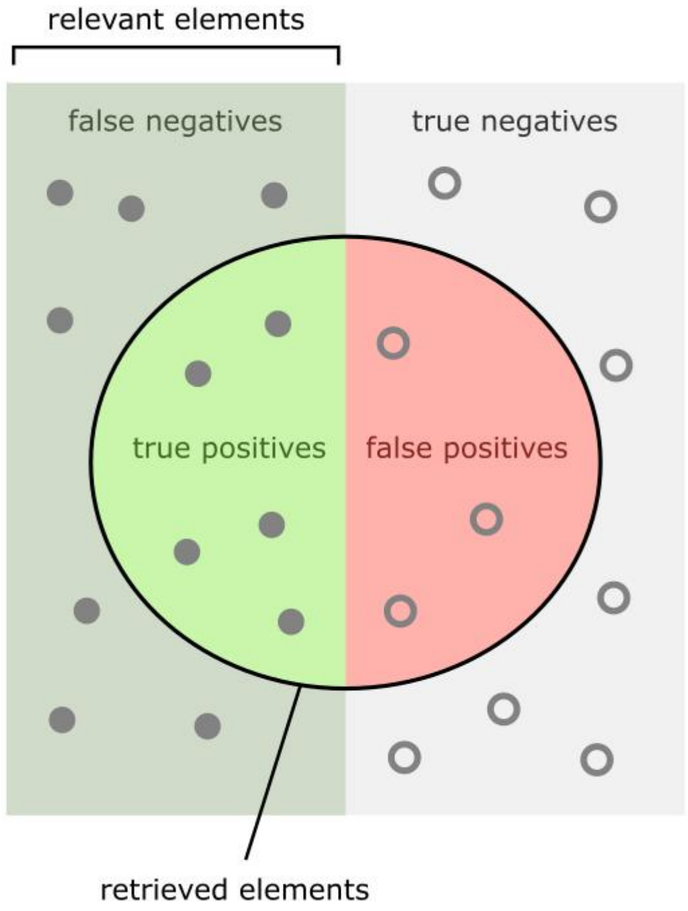
$$\text{Accuracy} = \frac{\text{correct predictions}}{\text{all predictions}} = \frac{52}{60}$$

Accuracy



$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Sensitivity *vs Specificity*

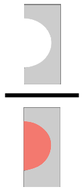


¿Cuántos elementos relevantes se seleccionan?

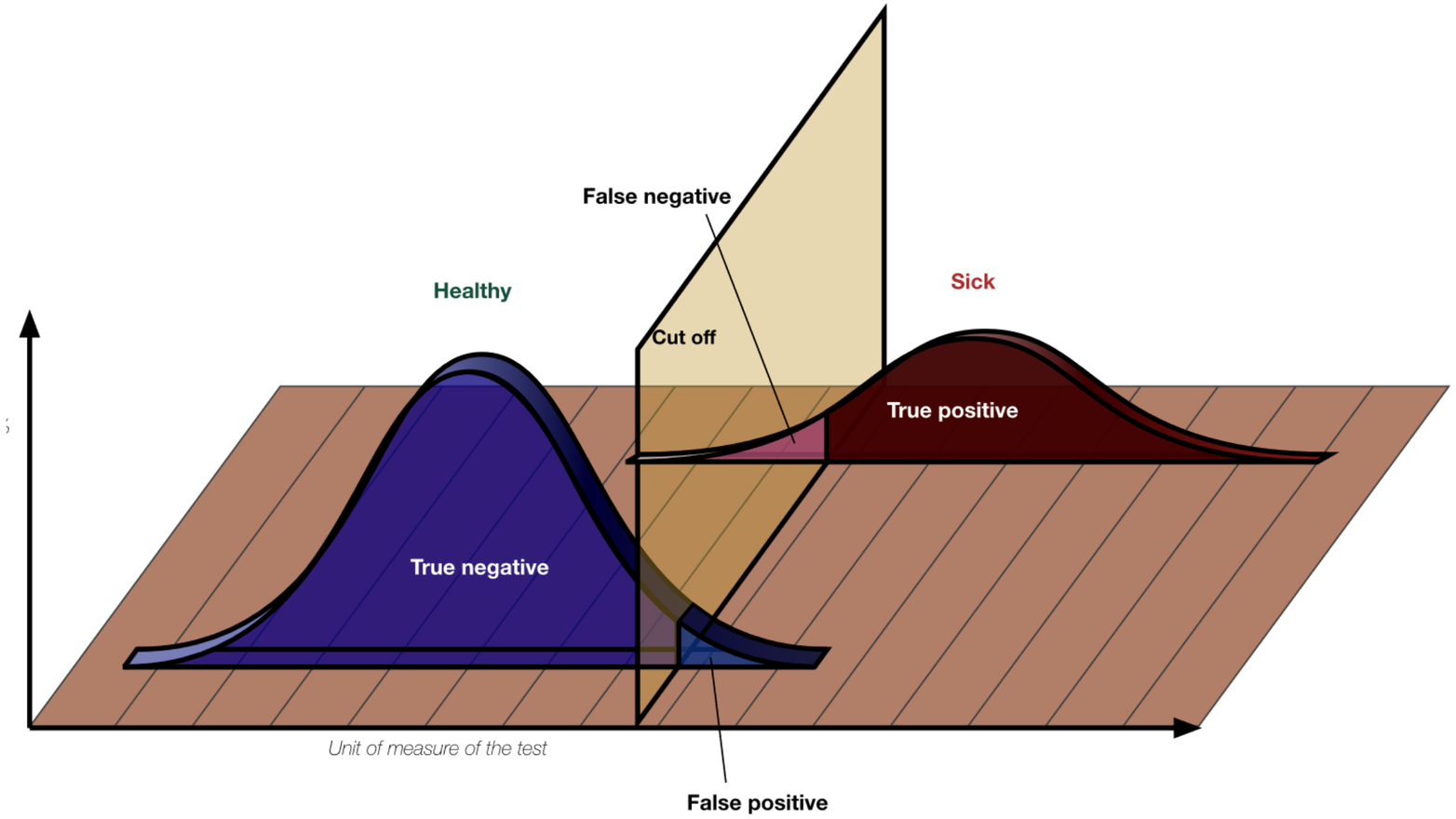
$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} =$$

¿Cuántos elementos negativos seleccionados son realmente negativos?

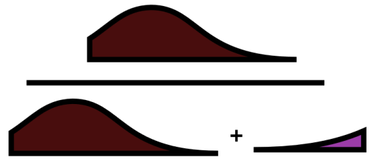
$$\text{Specificity} = \frac{\text{TN}}{\text{FP} + \text{TN}} =$$



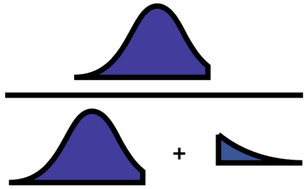
Sensitivity *vs Specificity*



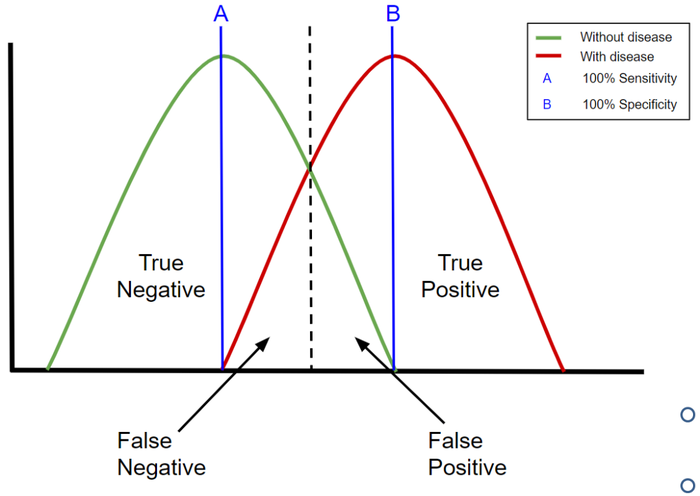
Sensitivity =



Specificity =

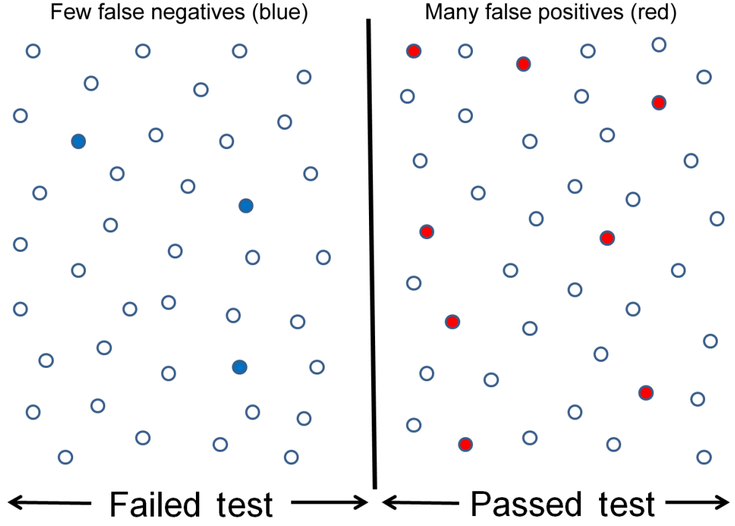


Sensitivity *vs Specificity*



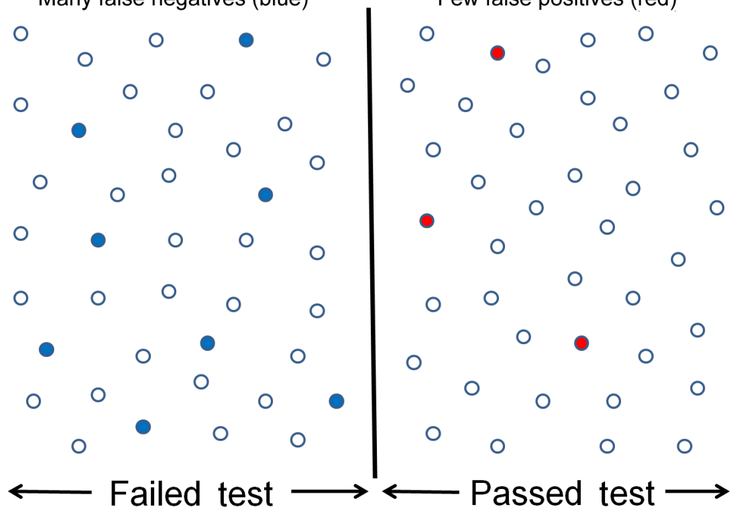
High sensitivity
Few false negatives (blue)

Low specificity
Many false positives (red)

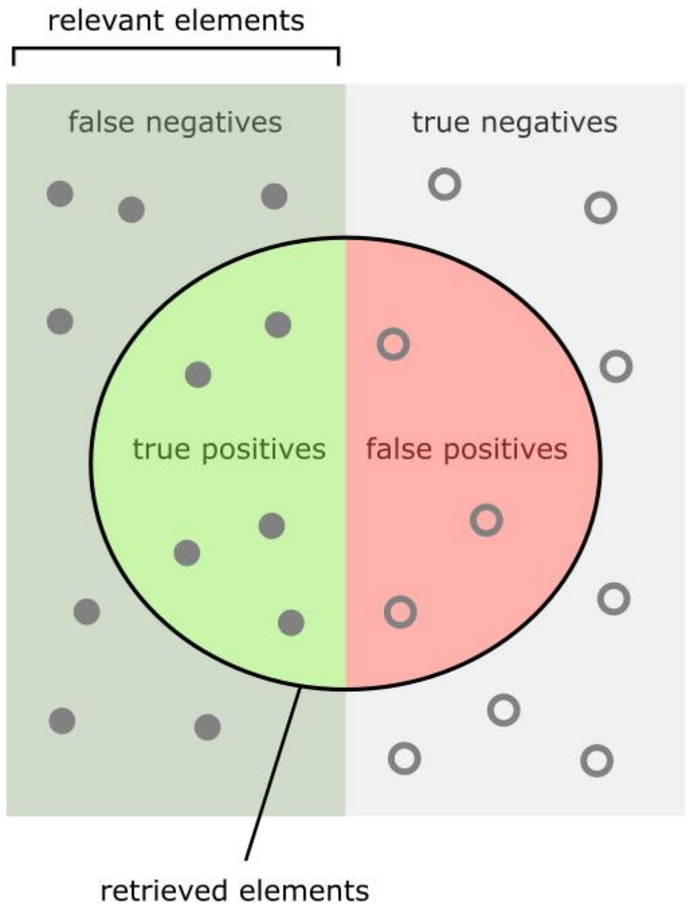


Low sensitivity
Many false negatives (blue)

High specificity
Few false positives (red)



Métricas



¿Cuántos elementos recuperados son relevantes?

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} =$$

¿Cuántos elementos relevantes se recuperan?

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} =$$

F1-score

El F1-score es una métrica de evaluación utilizada en problemas de clasificación para encontrar un balance entre la precision y el recall.

$$F1 = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

F1-score

F1-micro

- Calcula globalmente las verdaderas positivas (TP), falsas positivas (FP) y falsas negativas (FN), y luego aplica la fórmula del F1-score.
- Pondera por la frecuencia de cada clase y es útil cuando hay un desequilibrio de clases.

$$F1_{micro} = \frac{2 \cdot \sum TP}{2 \cdot \sum TP + \sum FP + \sum FN}$$

Se emplea cuando se necesita evaluar el rendimiento global, especialmente en datasets desbalanceados.

F1-score

F1-macro

- Calcula el F1-score para cada clase de manera individual y luego obtiene el promedio.
- Trata todas las clases por igual, sin importar su frecuencia.
- Útil si quieres evaluar el rendimiento en todas las clases de manera equitativa.

$$F1_{macro} = \frac{1}{n}\sum_{i=1}^{n}\frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}$$

Se emplea cuando quieres dar la misma importancia a cada clase, sin importar su frecuencia.

F1-score

F1-weighted

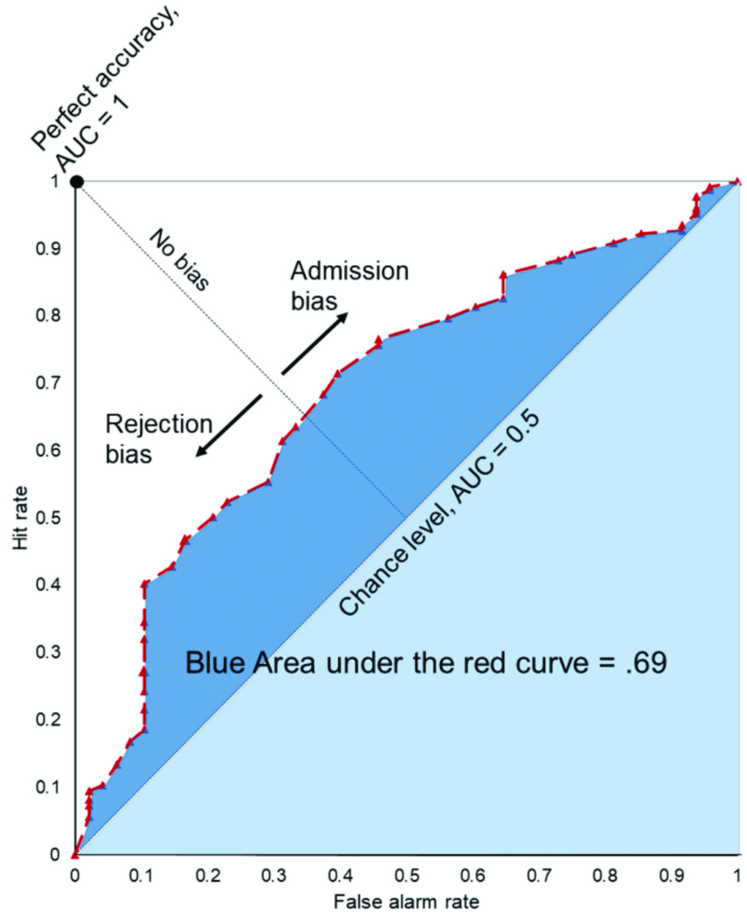
- Similar al F1-macro, pero pondera el F1-score de cada clase por el número de muestras de esa clase.
- Proporciona una métrica más representativa si el dataset está desbalanceado.

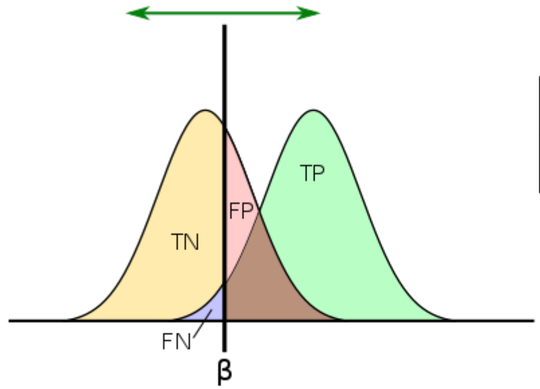
$$F1_{weighted} = \sum_{i=1}^{n} w_i \cdot F1_i$$

donde w_i es la proporción de la clase i.

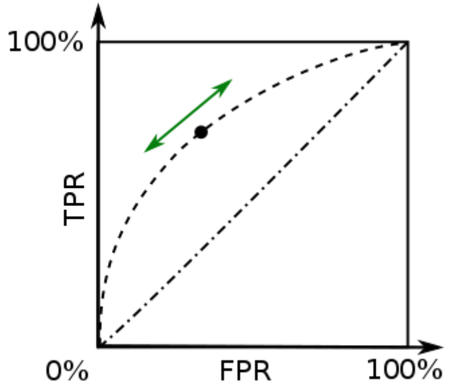
Se emplea cuando el desbalance de clases es significativo, pero aún quieres reflejar el tamaño de cada clase en la evaluación.

Receiver Operating *Characteristic curve (ROC)*





TP	FP
FN	TN



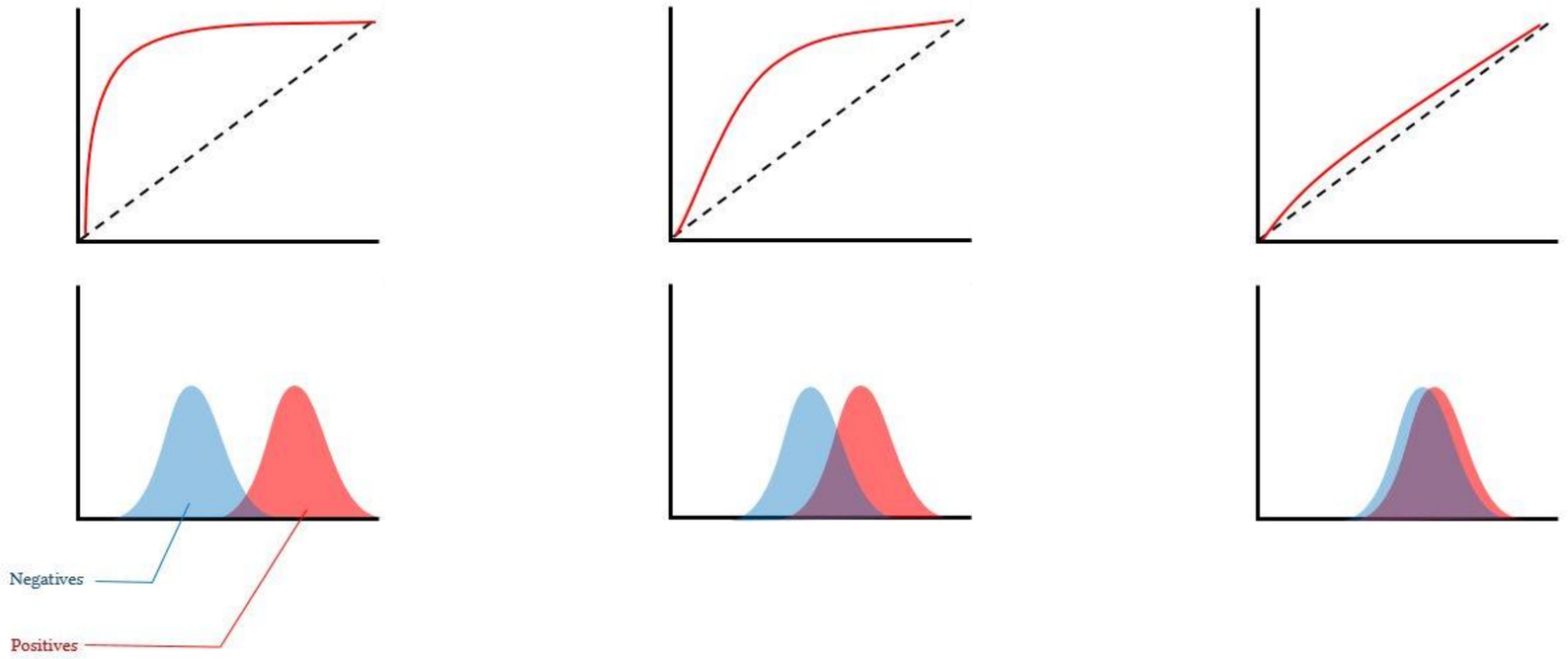
True Positive Rate (recall)

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$$

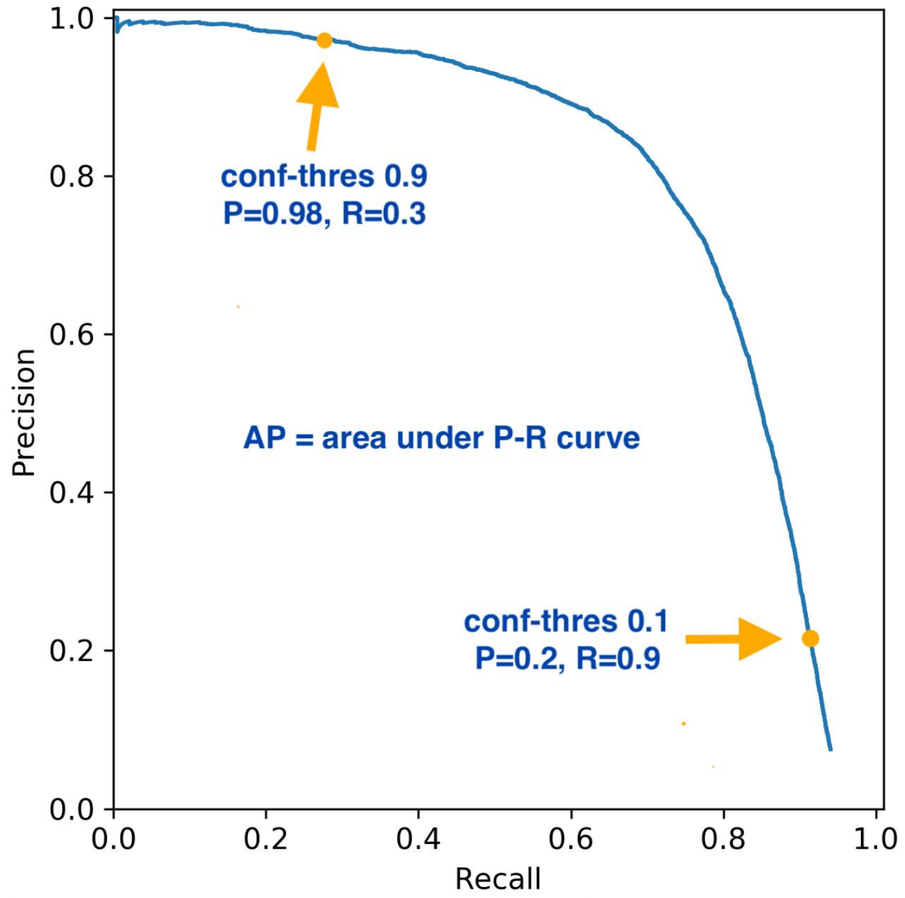
False Positive Rate

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}}$$

Receiver Operating *Characteristic curve (ROC)*



Precision-*Recall Curve*



$$\mathrm{AP} = \sum_{n=1}^{N} (\mathrm{R}_n - \mathrm{R}_{n-1})\mathrm{P}_k$$

donde P_n y R_n son precision y recall en el n-ésimo threshold.
Al par $(\mathrm{P}_n, \mathrm{R}_n)$ es conocido como operating point.

Métricas

	Receiver Operating Characteristic curve	Precision-Recall Curve
	Muestra la habilidad general de clasificación.	Muestra el rendimiento del modelo sobre la clase positiva.
Cuando usar	• Cuando las clases son balanceadas (o aproximadamente balanceadas). • Cuando se desea evaluar la discriminación global del modelo.	• Cuando los datos son altamente desbalanceados. • Cuando el interés principal es reducir los falsos positivos.

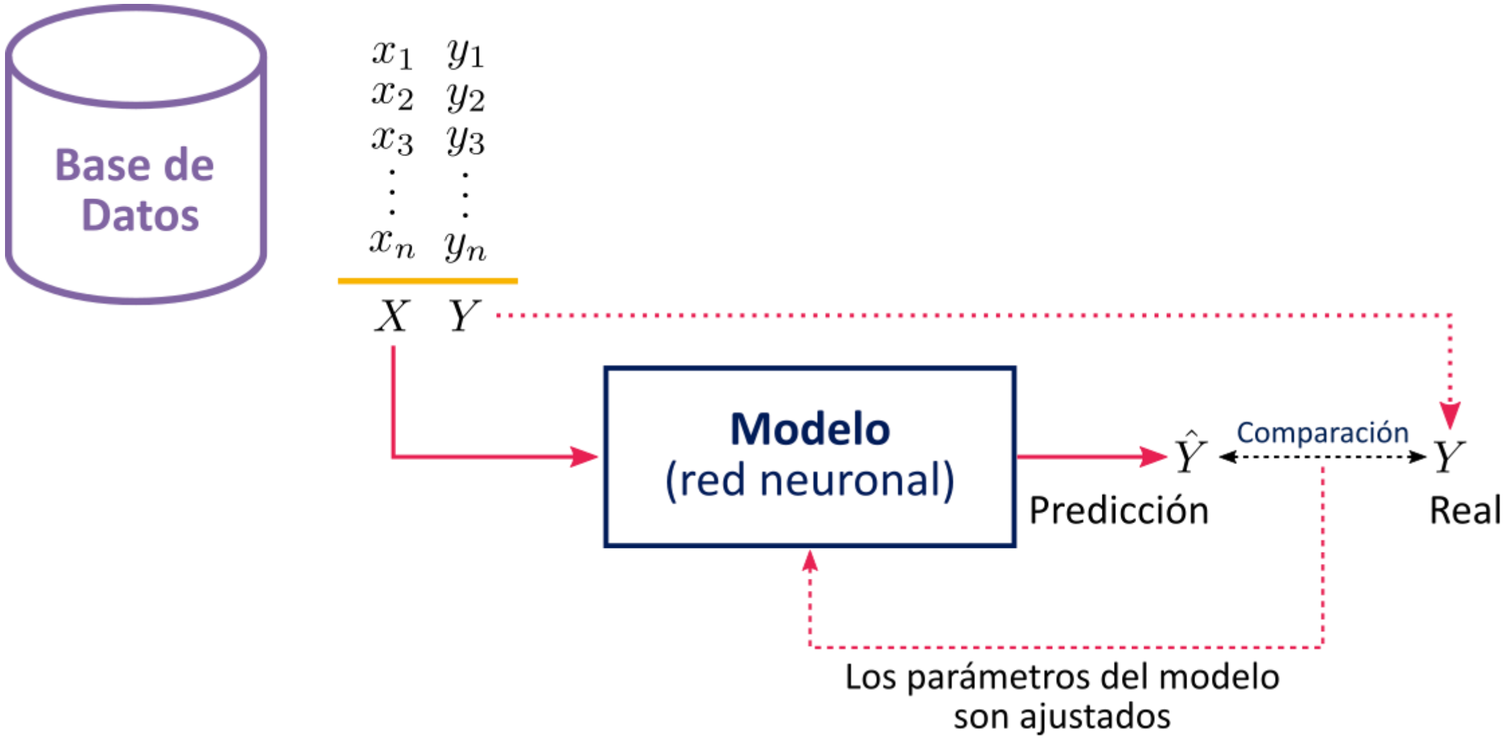
UTEC
UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

2.

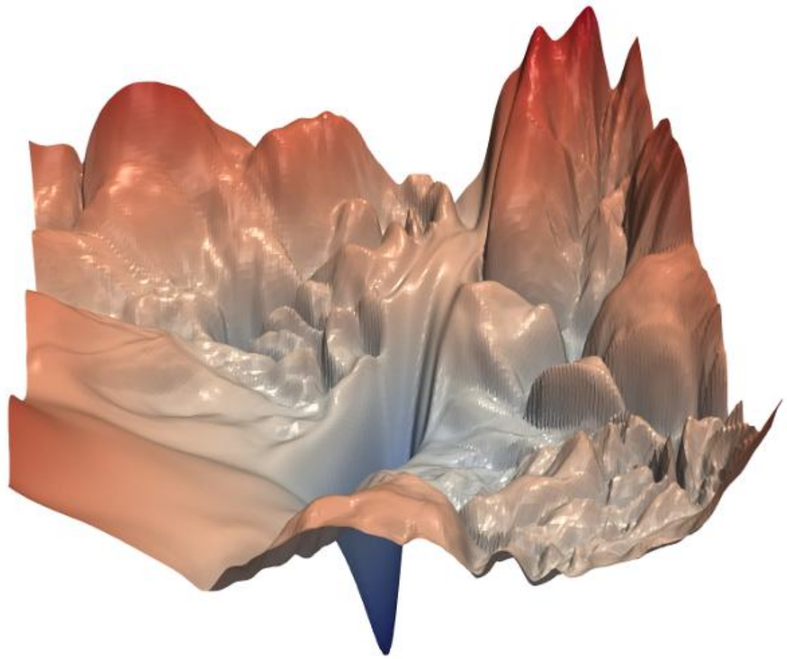
Training *tricks*

TRANSFORMATEC

Training


Base de Datos
x_1 y_1
x_2 y_2
x_3 y_3
x_n y_n
X Y
Modelo
(red neuronal)
$\hat{Y}$
Comparación
Y
Predicción
Real
Los parámetros del modelo son ajustados

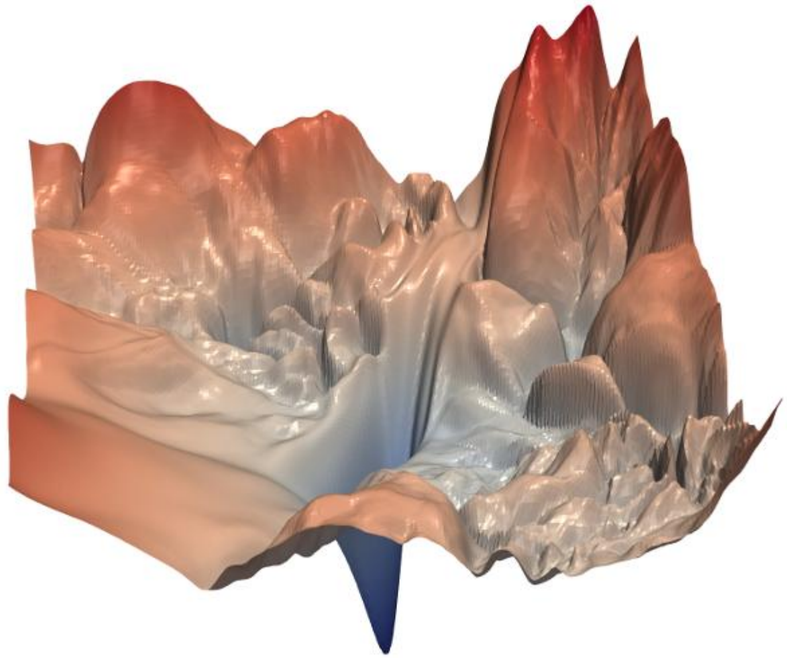
Loss *landscape*



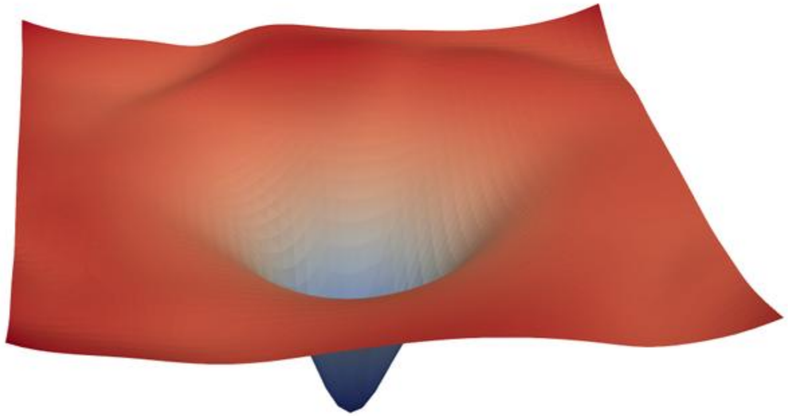
(a) without skip connections

Hao Li et al. (2017) "Visualizing the Loss Landscape of Neural Nets". arXiv e-prints. arXiv preprint arXiv:1712.09913.

Loss *landscape*



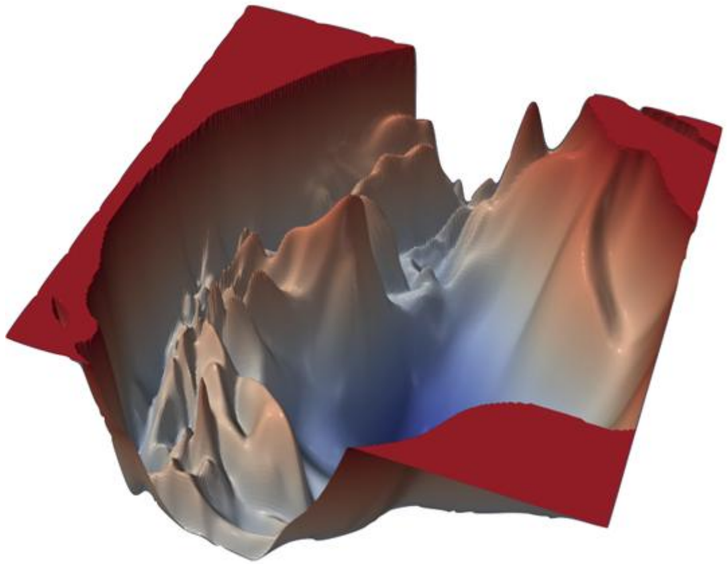
(a) without skip connections



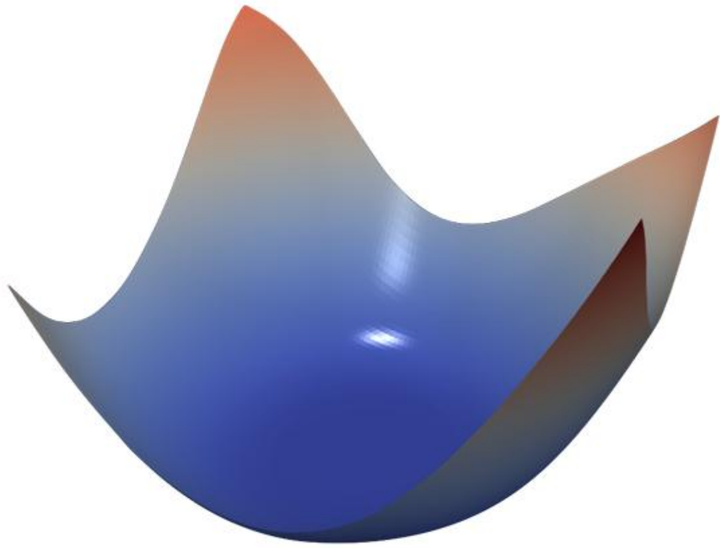
(b) with skip connections

Hao Li et al. (2017) "Visualizing the Loss Landscape of Neural Nets". arXiv e-prints. arXiv preprint arXiv:1712.09913.

Loss *landscape*



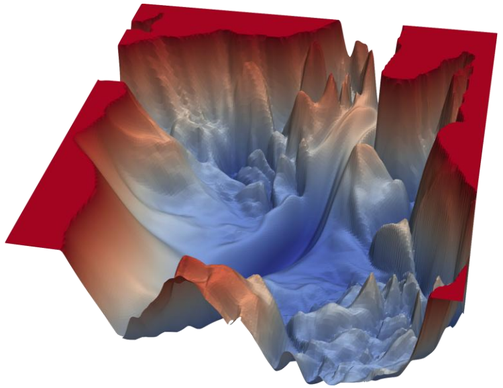
(a) ResNet-110, no skip connections



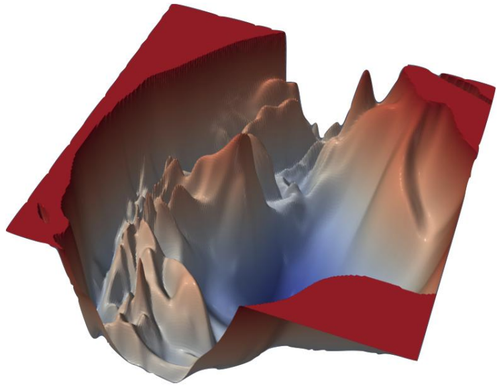
(b) DenseNet, 121 layers

Hao Li et al. (2017) "Visualizing the Loss Landscape of Neural Nets". arXiv e-prints. arXiv preprint arXiv:1712.09913.

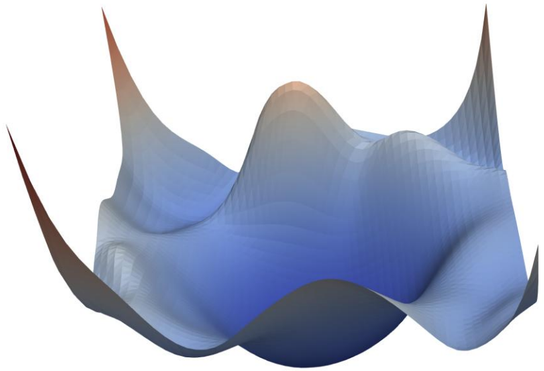
Loss *landscape*



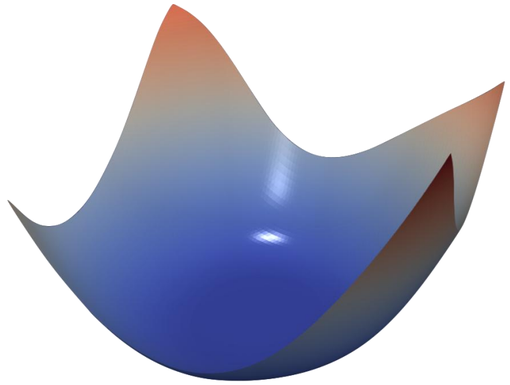
VGG-56



ResNet-110 without skip connections



Renset-56

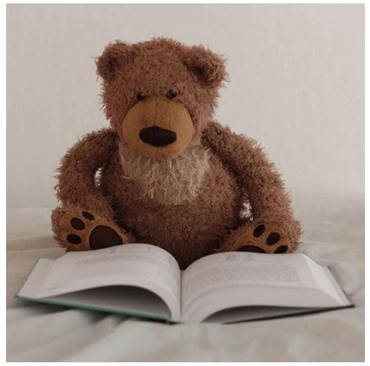


Densenet-121

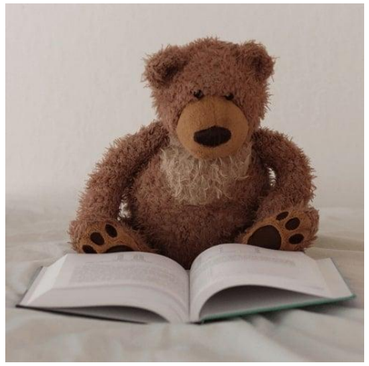
Hao Li et al. (2017) "Visualizing the Loss Landscape of Neural Nets". arXiv e-prints. arXiv preprint arXiv:1712.09913.

Data *augmentation*

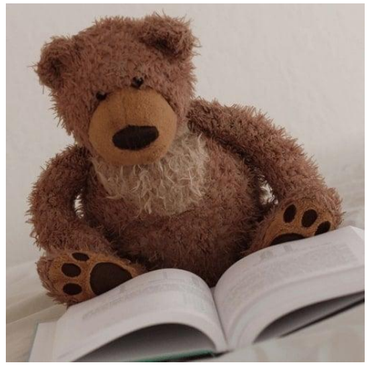
Original



Flip

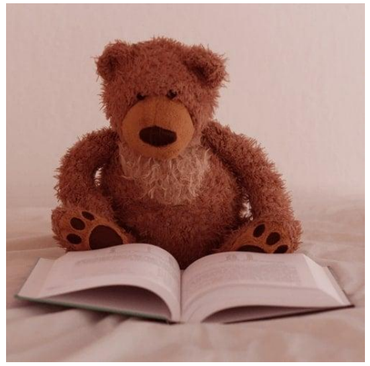


Rotation



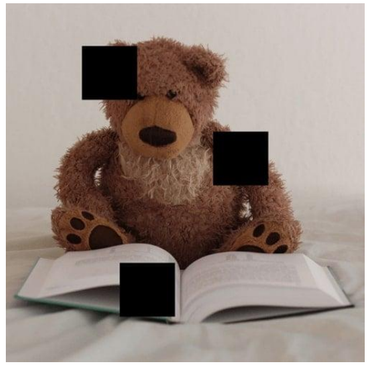
Random crop

Color shift

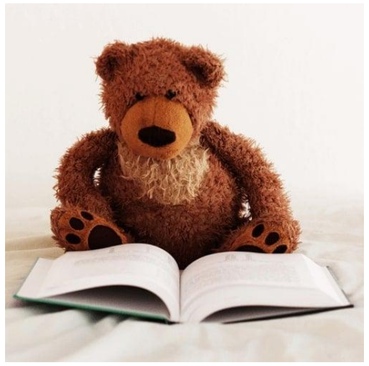


Noise addition

Information loss

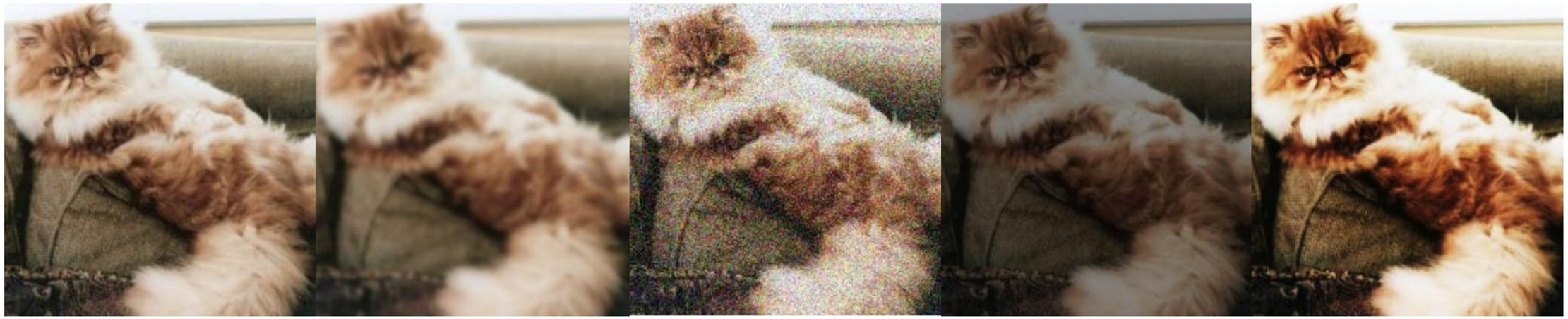


Contrast change



Connor Shorten and Taghi M. Khoshgoftaar (2019) "A survey on Image Data Augmentation for Deep Learning". Journal of Big Data

Data *augmentation*



Original **Blur** **Noise** **Brillo** **Contraste**

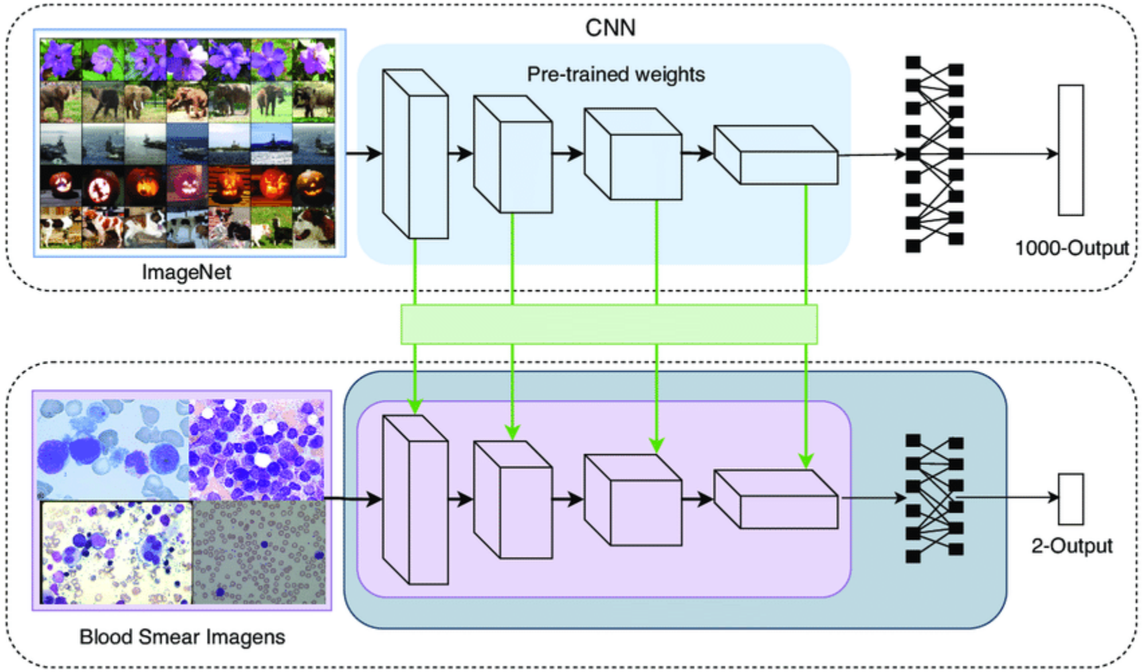
Connor Shorten and Taghi M. Khoshgoftaar (2019) "A survey on Image Data Augmentation for Deep Learning". Journal of Big Data

Data *augmentation*

Día/Noche

Nevado/Soleado

Pre-*training*



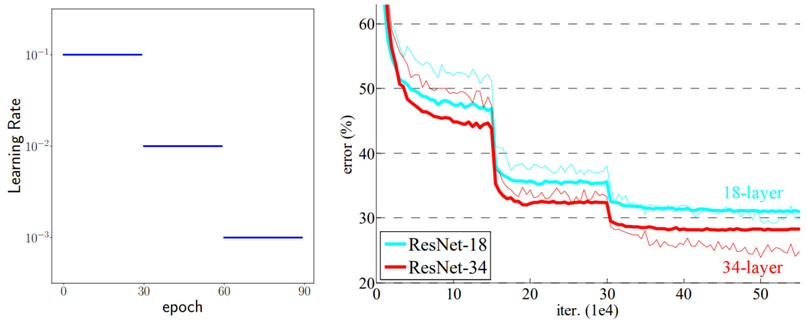
Pre-*training*

	ImageNet domain	Not similar to ImageNet
Small Dataset	Transfer learning	Recopilar más datos
Big Dataset	Fine-tuning	Entrenar desde cero

Learning *rate decay*

Step LR $\alpha = \alpha_0 \gamma^{\lfloor \frac{n}{s} \rfloor}$

- n: época.
- s: Período de caída de la tasa de aprendizaje.
- γ:Factor multiplicativo del decaimiento de la tasa de aprendizaje.
- α_0: taza de aprendizaje inicial.



(a) Learning rate decay strategy (b) Figure taken from He et al. (2016)

Figure 1: Training error in (b) is shown by thin curves, while test error in (b) by bold curves.

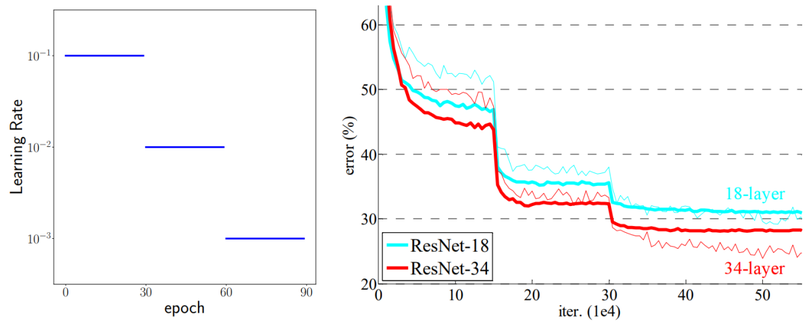
Learning *rate decay*

Step LR $\alpha = \alpha_0 \gamma^{\lfloor \frac{n}{s} \rfloor}$

- n: época.
- s: Período de caída de la tasa de aprendizaje.
- γ:Factor multiplicativo del decaimiento de la tasa de aprendizaje.
- α_0: taza de aprendizaje inicial.

Decaimiento exponencial $\alpha = \alpha_0 e^{-kn}$

Decaimiento 1/t $\alpha = \alpha_0 \frac{1}{1+kn}$



(a) Learning rate decay strategy (b) Figure taken from He et al. (2016)

Figure 1: Training error in (b) is shown by thin curves, while test error in (b) by bold curves.

Learning *rate decay*

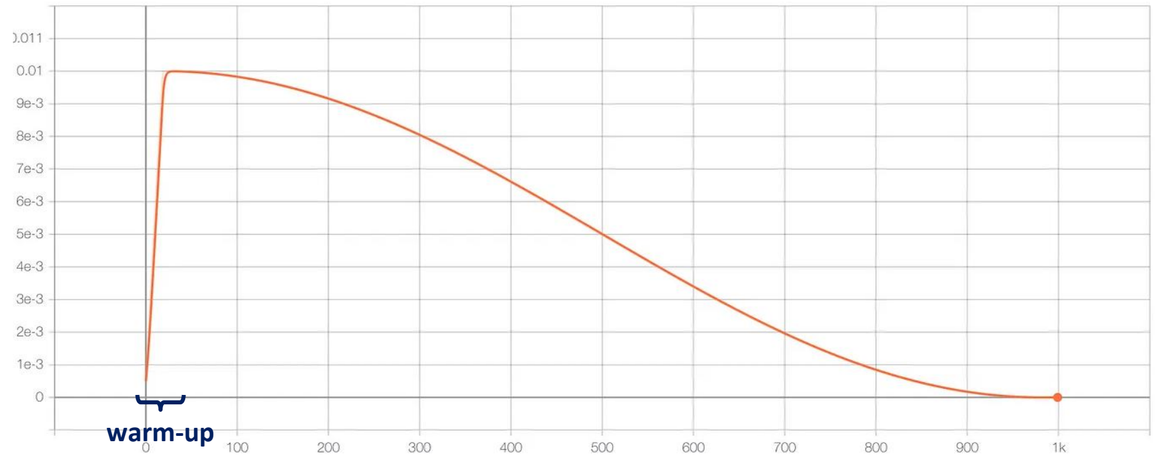
Step LR $\alpha = \alpha_0 \gamma^{\lfloor \frac{n}{s} \rfloor}$

- n: época.
- s: Período de caída de la tasa de aprendizaje.
- γ:Factor multiplicativo del decaimiento de la tasa de aprendizaje.
- α_0: taza de aprendizaje inicial.

Decaimiento exponencial $\alpha = \alpha_0 e^{-kn}$

Decaimiento 1/t $\alpha = \alpha_0 \frac{1}{1 + kn}$

Decaimiento coseno $\alpha = \alpha_0 \cdot \frac{1}{2}\left(1 + \cos\left(\frac{n}{N}\pi\right)\right)$



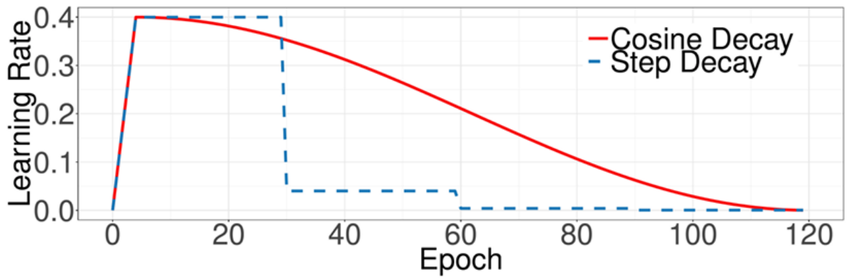
Bag of *Tricks*

Cosine Learning Rate Decay

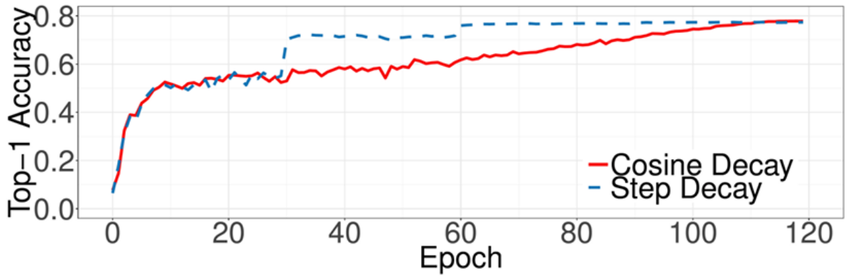
$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min})\left(1 + \cos\left(k\frac{\pi}{\tau}\right)\right)\eta$$

- η_t es la tasa de aprendizaje en el paso t.
- $\eta_{\min}$ es el valor mínimo de la tasa de aprendizaje.
- $\eta_{\max}$ es el valor máximo de la tasa de aprendizaje.
- τ es el número total de *epochs*.
- k es el número de *epochs* transcurridos hasta el momento actual.

Learning Rate Schedule



Validation Accuracy



Tong He et al. (2018) "Bag of Tricks for Image Classification with Convolutional Neural Networks". Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 558-567).

Bag of *Tricks*

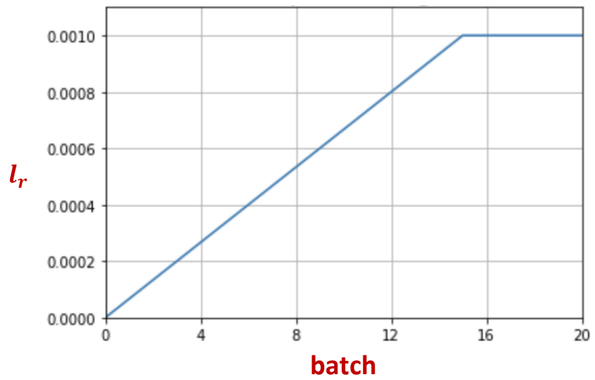
Large-batch training

Linear scaling learning rate

Se propone aumentar la tasa de aprendizaje de manera proporcional al tamaño del *batch*.

$$l_r = 0.1\frac{b}{256}$$

Learning rate warmup



Zero γ

El parámetro de escala γ se inicializa en 0:

$$\mathrm{BN}_{\gamma,\beta}(z_i) = \boldsymbol{\gamma}\hat{z}_i + \boldsymbol{\beta}$$

Esto hace que, al inicio del entrenamiento, la salida de esos bloques sea esencialmente la misma que la entrada (es decir, los bloques no aportan cambios significativos a los datos que procesan).

No bias decay

Weight decay: $g_t = \nabla_\theta \mathcal{L}(\theta_t) + \lambda\theta_t$

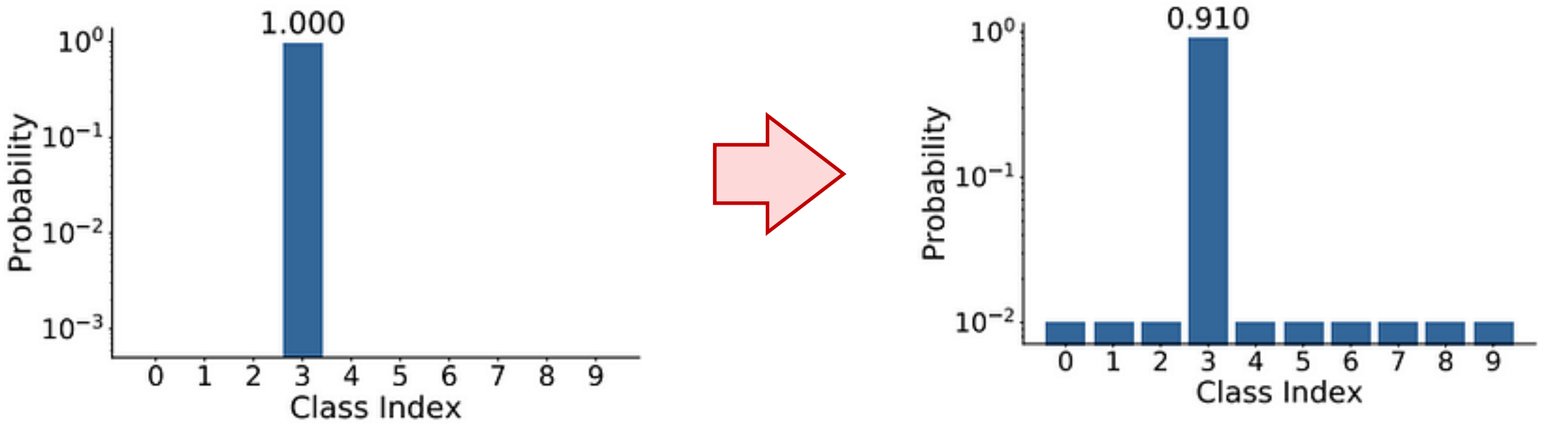
Se recomienda aplicar el *weight decay* solo a los pesos de las capas convolucionales y *fully-connected*, excluyendo los *biases* y los parámetros γ y β de las capas BN. Esto reduce el riesgo de sobreajuste y mejora la generalización.

Tong He et al. (2018) "Bag of Tricks for Image Classification with Convolutional Neural Networks". Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 558-567).

Bag of *Tricks*

Label Smoothing

$$p_i = \begin{cases} 1-\varepsilon, & i = y \\ \dfrac{\varepsilon}{K-1}, & i \neq y \end{cases}$$

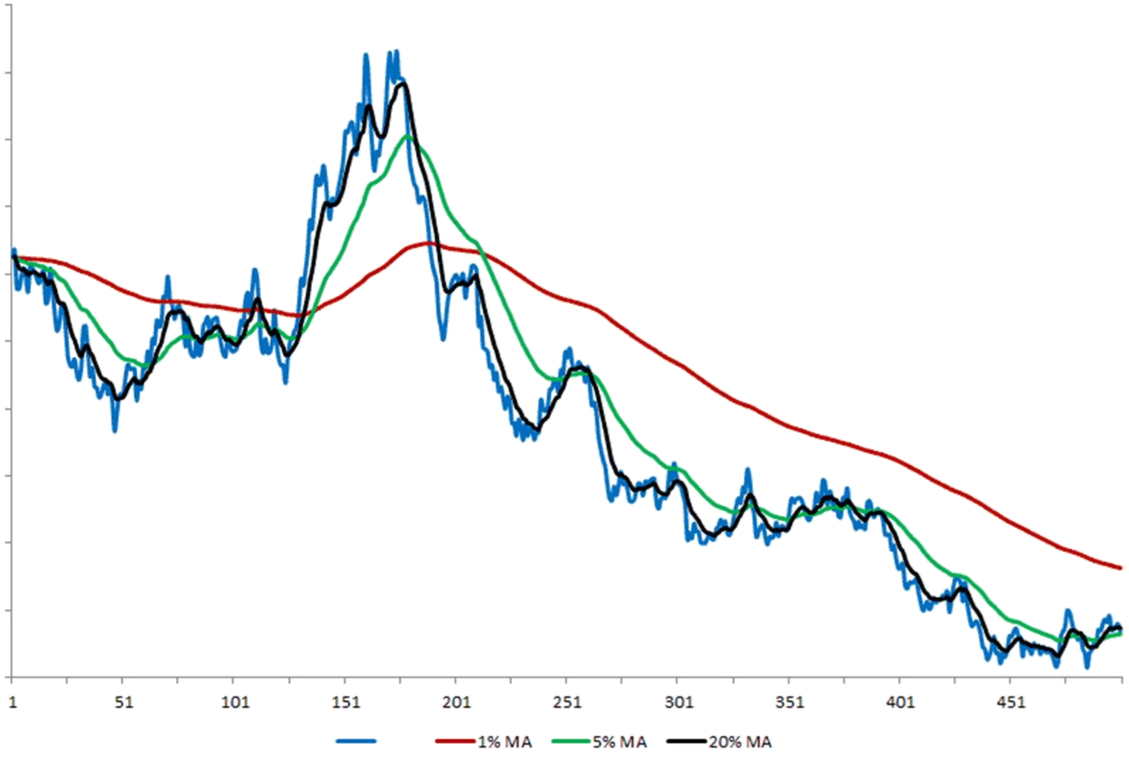


Tong He et al. (2018) "Bag of Tricks for Image Classification with Convolutional Neural Networks". Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 558-567).

ResNet-RS

(Re-Scaling ResNet)

Exponential moving average (EMA) of the weights



$$\theta_t^{\mathrm{EMA}} = \alpha\theta_{t-1}^{\mathrm{EMA}} + (1-\alpha)\theta_t$$

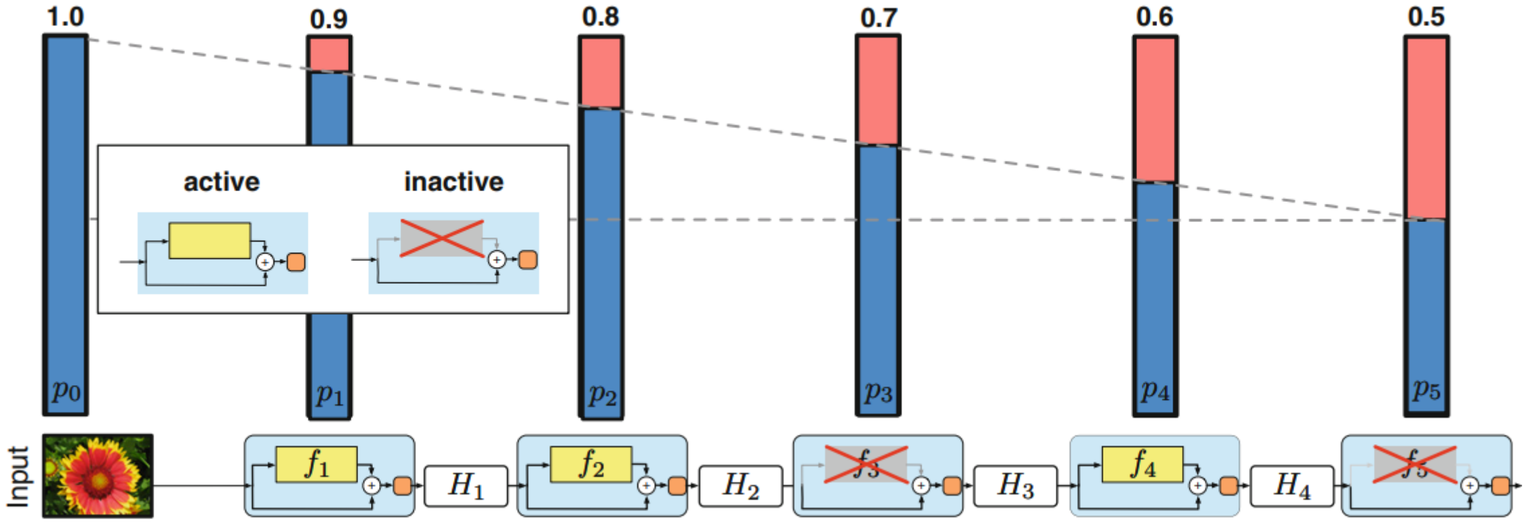
- θ_t^{EMA} es el valor del parámetro suavizado en el paso t.
- α es el factor de suavizado, que controla la velocidad con la que se "olvidan" los valores antiguos. Típicamente, α se establece en un valor cercano a 1, como 0.999.
- θ_t es el valor actual del parámetro en el paso t.

Boris T. Polyak et al. (1992) "Acceleration of stochastic approximation by averaging". SIAM journal on control and optimization, 30(4), 838-855.

ResNet-RS

(Re-Scaling ResNet)

Stochastic Depth

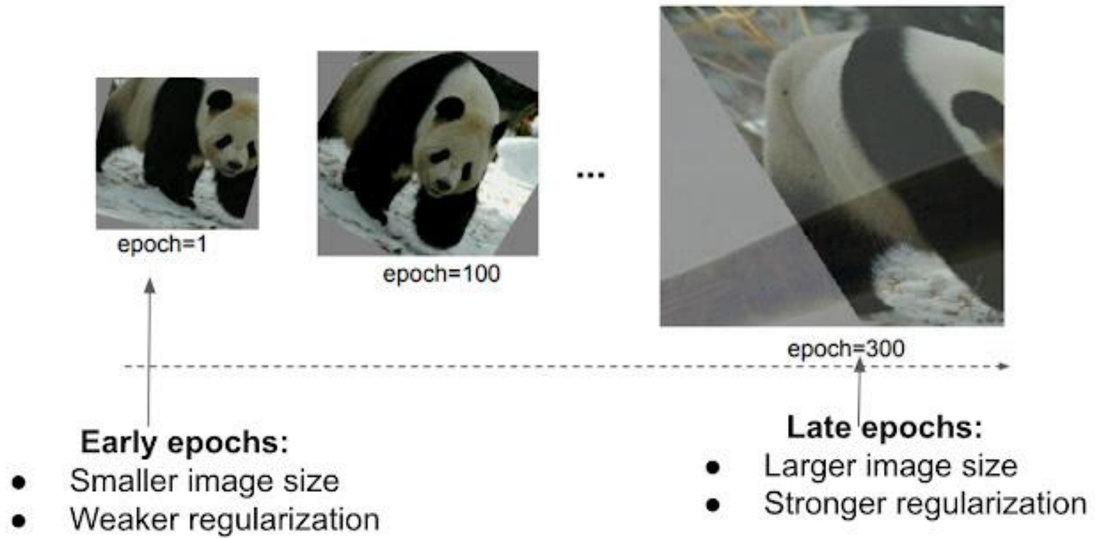


$$p_\ell = 1 - \frac{\ell}{L}(1 - p_L)$$

Gao Huang et al. (2016) "Deep networks with stochastic depth". Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14 (pp. 646-661). Springer International Publishing.

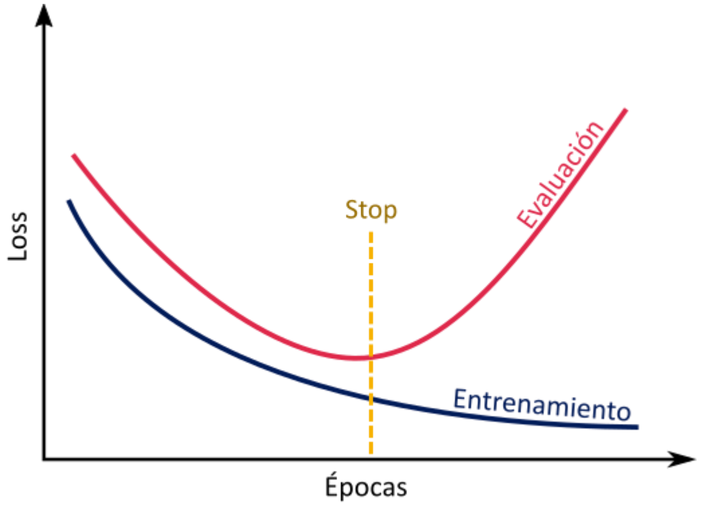
EfficientNetV2

Progressive Learning



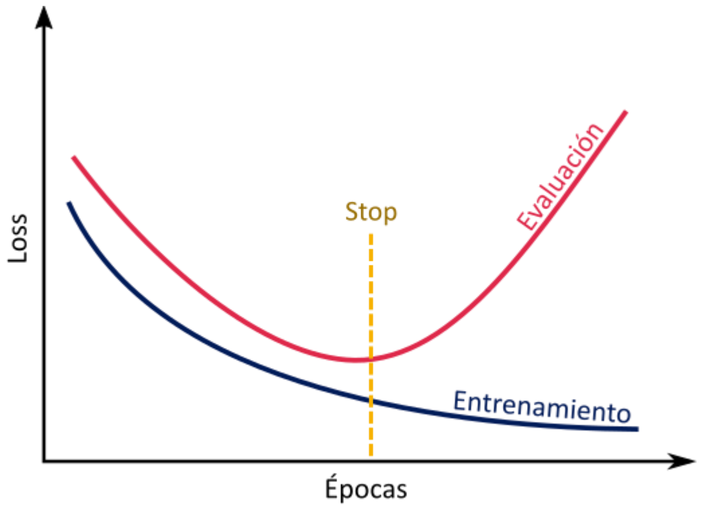
Mingxing Tan et al. (2021) "EfficientNetV2: Smaller Models and Faster Training". arXiv preprint arXiv:2104.00298.

Early *stopping*



Lutz Prechelt (2022) "Early Stopping — But When?"
Springer Berlin Heidelberg, 2002. 55-69.

Early *stopping*



Valor de Loss de entrenamiento: $\mathcal{L}_{tr}$

Valor de Loss de evaluación: $\mathcal{L}_{ev}$

Mejor valor obtenido de $\mathcal{L}_{ev}$:

$$\mathcal{L}_{op} = \min_{t' \leq t} \mathcal{L}_{ev}(t')$$

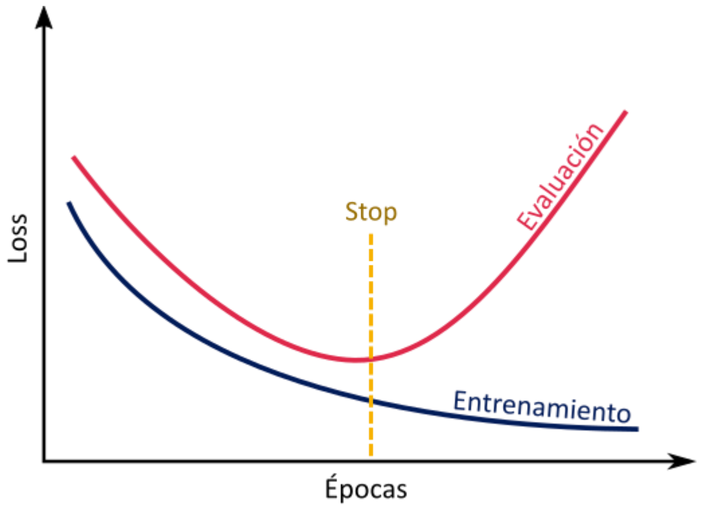
Criterio 1:

$$\mathrm{GL}(t) = 100\left(\frac{\mathcal{L}_{ev}(t)}{\mathcal{L}_{op}(t)} - 1\right)$$

GL_α: Parada después de la primera época t con $\mathrm{GL}(t) > \alpha$

Lutz Prechelt (2022) "Early Stopping — But When?"
Springer Berlin Heidelberg, 2002. 55-69.

Early *stopping*



Valor de Loss de entrenamiento: $\mathcal{L}_{tr}$

Valor de Loss de evaluación: $\mathcal{L}_{ev}$

Mejor valor obtenido de $\mathcal{L}_{ev}$:

$$\mathcal{L}_{op} = \min_{t' \le t} \mathcal{L}_{ev}(t')$$

Criterio 1:

$$\mathrm{GL}(t) = 100\left(\frac{\mathcal{L}_{ev}(t)}{\mathcal{L}_{op}(t)} - 1\right)$$

GL_α: Parada después de la primera época t con $\mathrm{GL}(t) > \alpha$

Criterio 2:

$$\mathrm{P}_k(t) = 1000\left(\frac{\mu_k(t)}{\delta_k(t)} - 1\right) \quad \text{donde:} \quad \mu_k(t) = \frac{1}{k}\sum_{t'=t-k+1}^{t} \mathcal{L}_{tr}(t'),$$

$$\delta_k(t) = \min_{t' \in [t-k+1,t]} \mathcal{L}_{tr}(t')$$

PQ_α: Parada después de la primera época de final de franja t con $\frac{\mathrm{GL}(t)}{\mathrm{P}_k(t)} > \alpha$

Lutz Prechelt (2022) "Early Stopping — But When?"
Springer Berlin Heidelberg, 2002. 55-69.

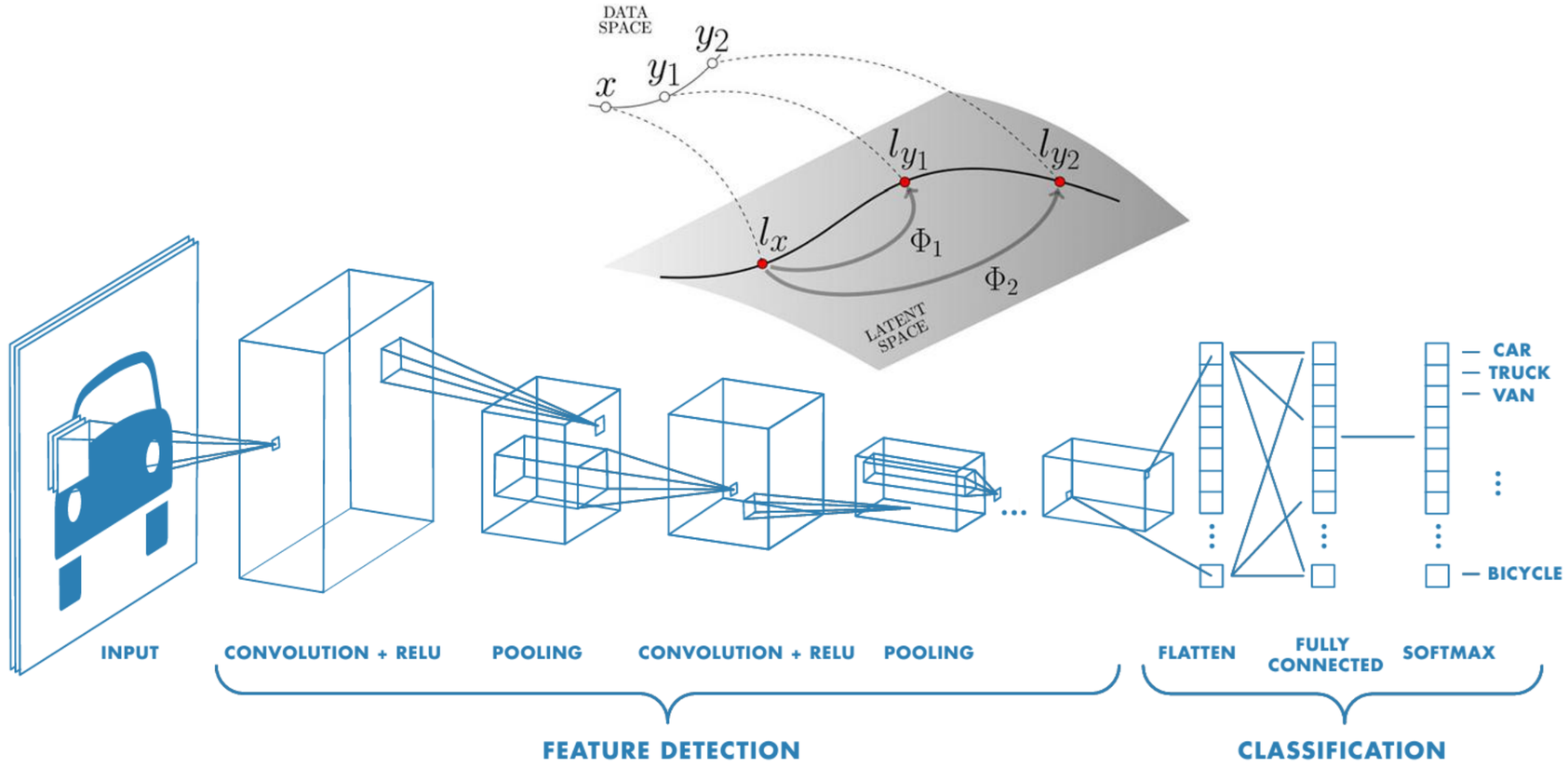
UTEC
UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

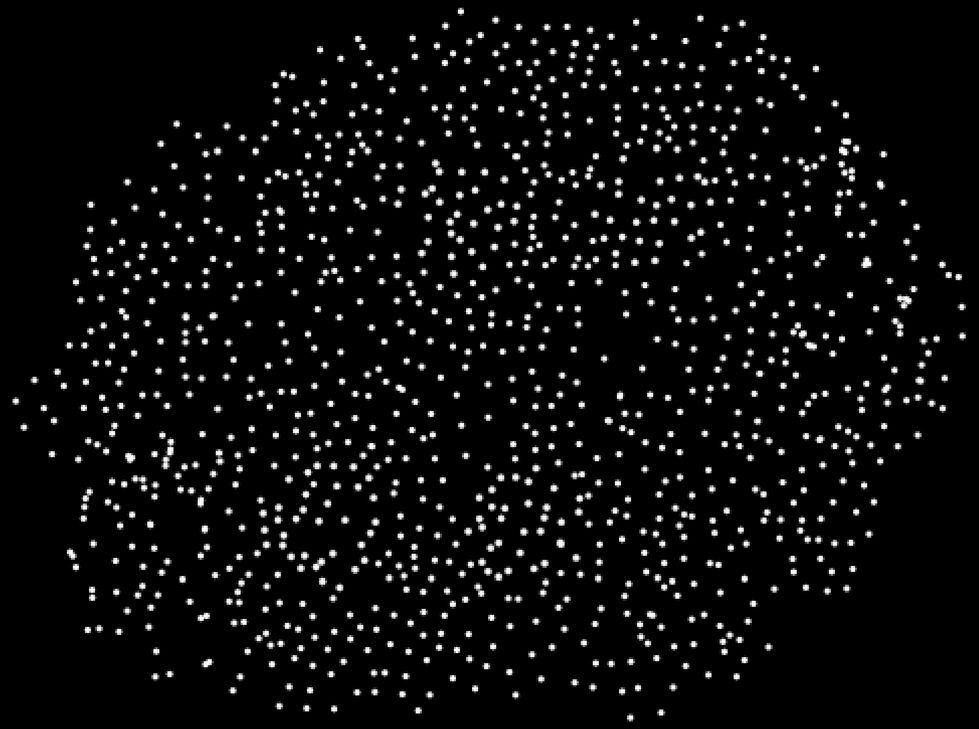
3.

Learning *representation*

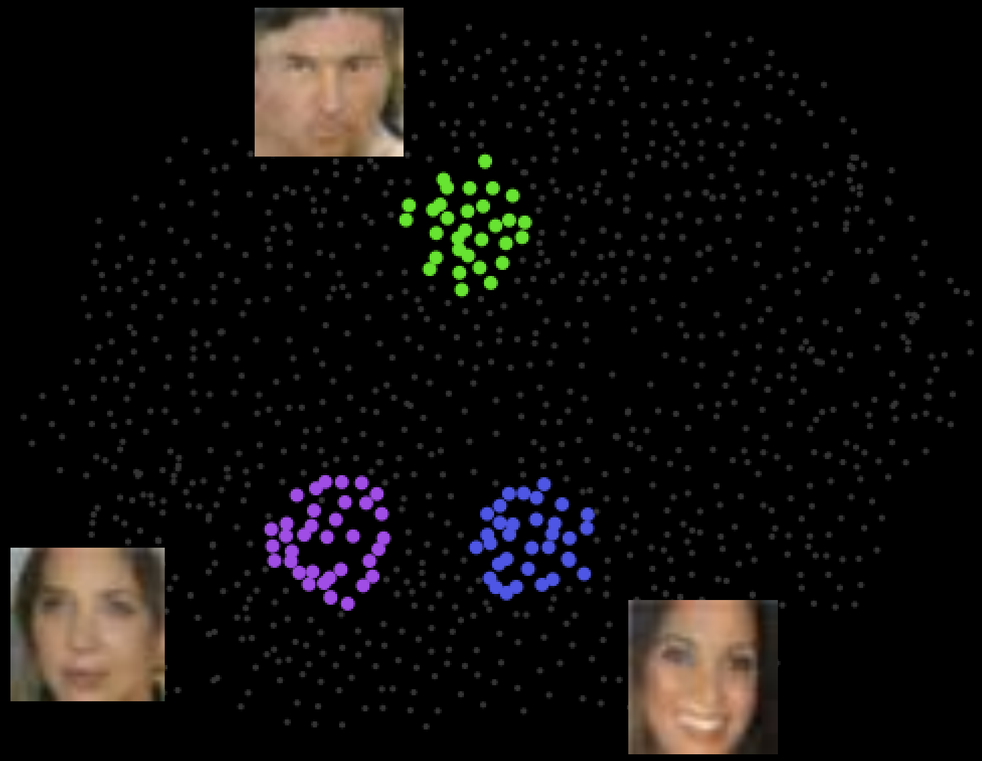
TRANSFORMATEC

Deep learning *pipeline*



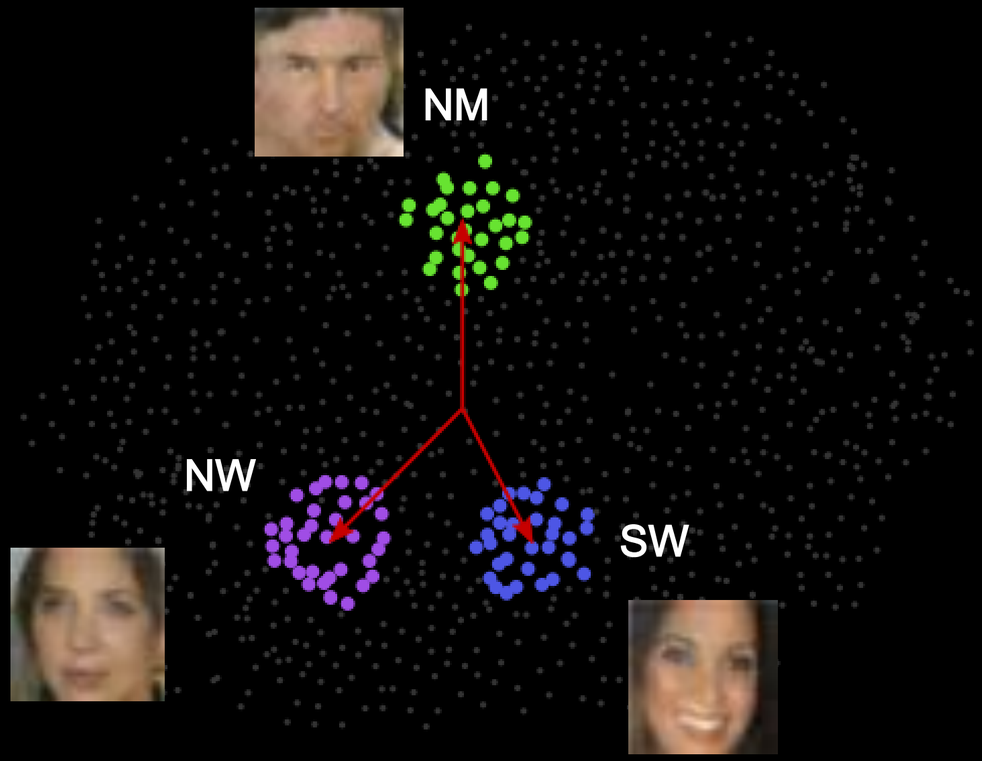


Latent space



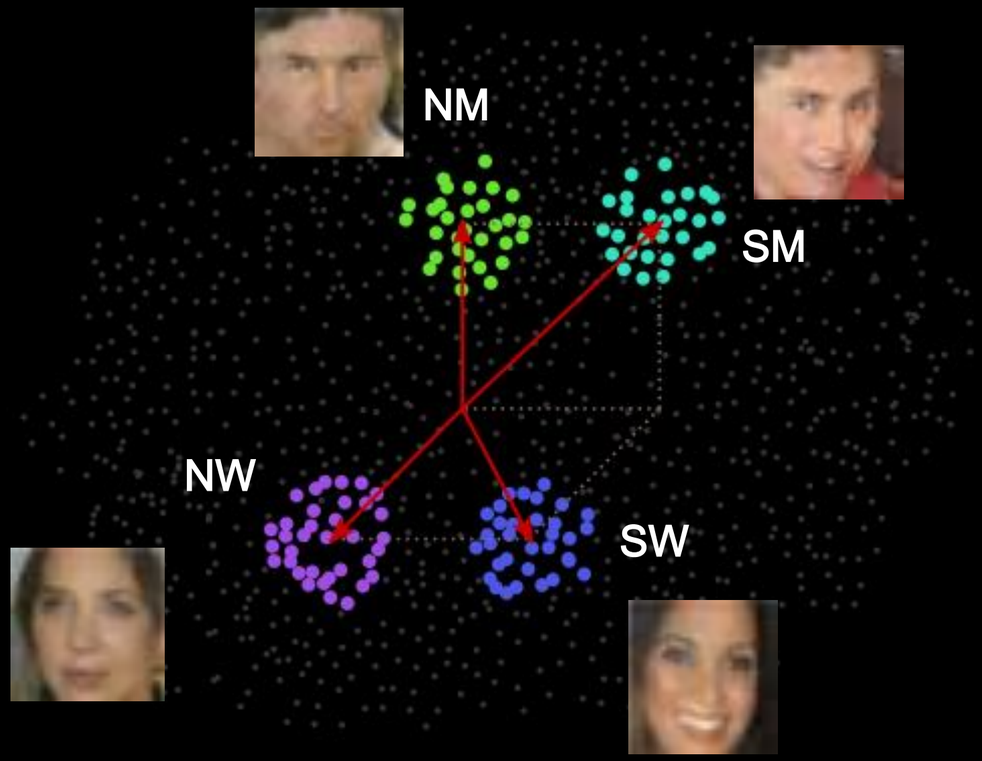
Latent space

Alec Radford et al. (2016) “Unsupervised representation learning with deep convolutional generative adversarial networks”.
4th International Conference on Learning Representations (ICRL) , Conference Track Proceedings.



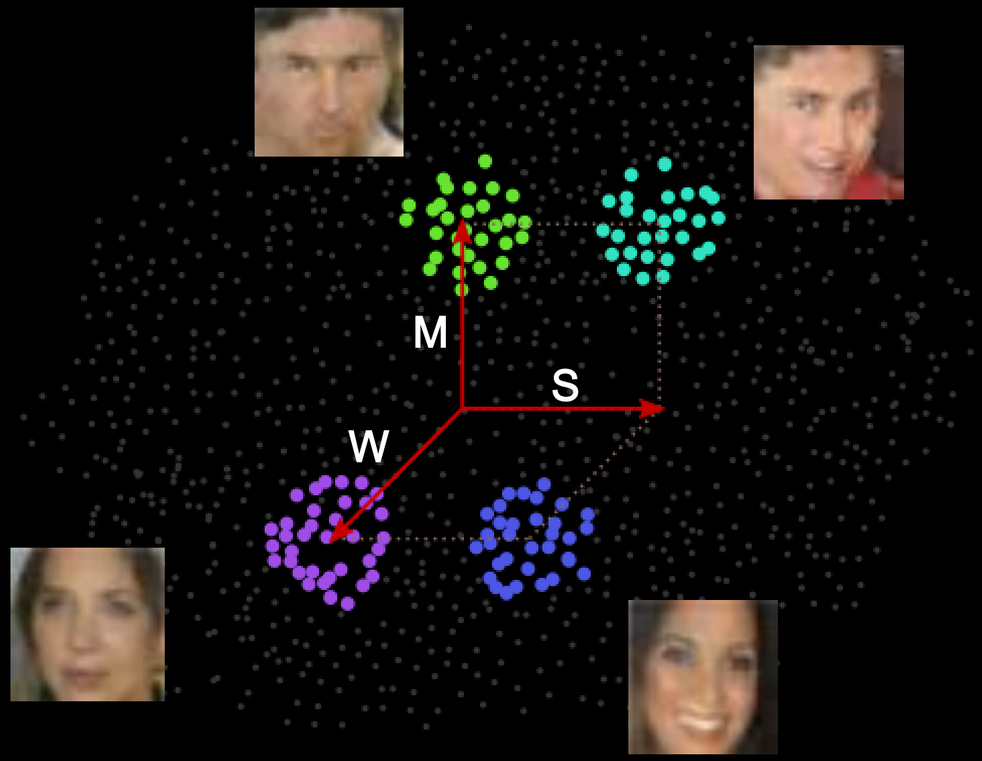
Latent space

Alec Radford et al. (2016) “Unsupervised representation learning with deep convolutional generative adversarial networks”.
4th International Conference on Learning Representations (ICRL) , Conference Track Proceedings.



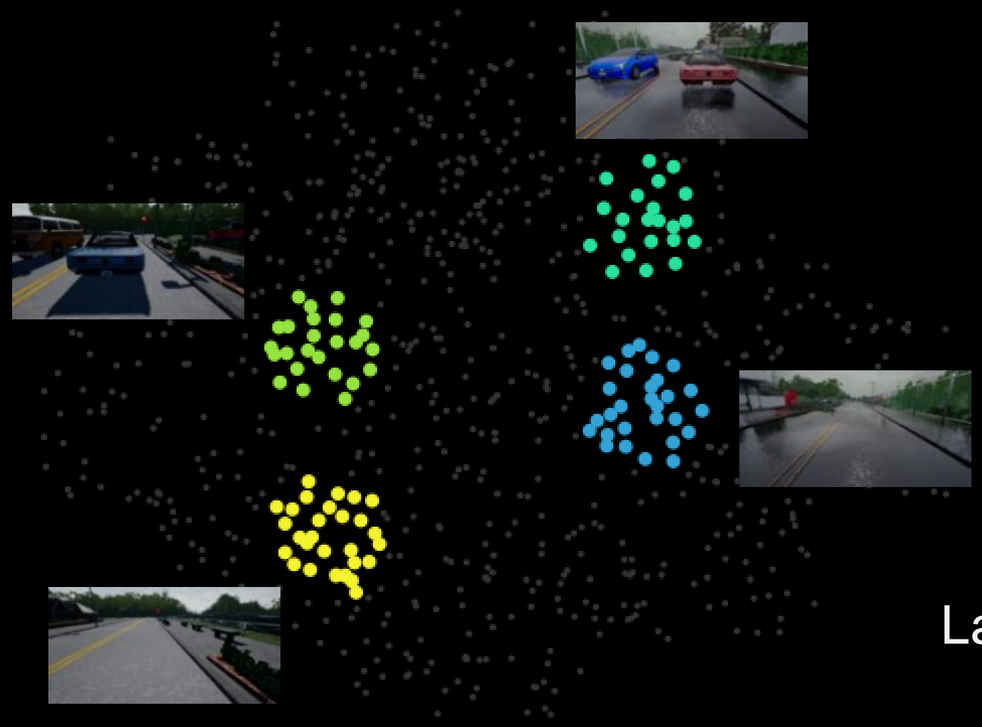
Latent space

Alec Radford et al. (2016) “Unsupervised representation learning with deep convolutional generative adversarial networks”.
4th International Conference on Learning Representations (ICRL) , Conference Track Proceedings.

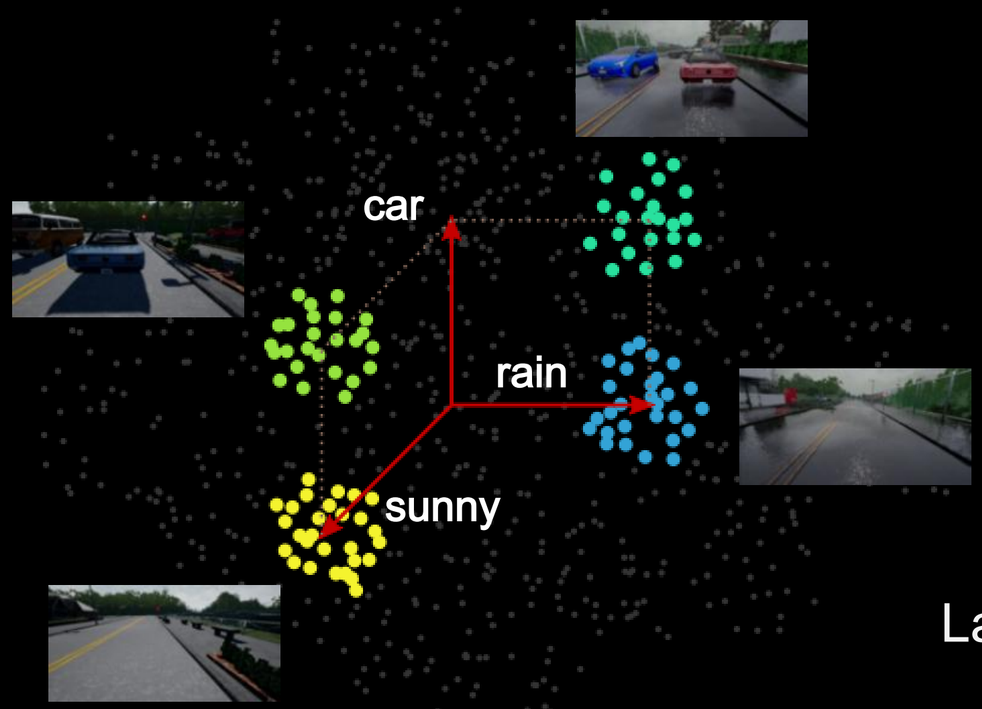


Latent space

Alec Radford et al. (2016) “Unsupervised representation learning with deep convolutional generative adversarial networks”.
4th International Conference on Learning Representations (ICRL) , Conference Track Proceedings.



Latent space

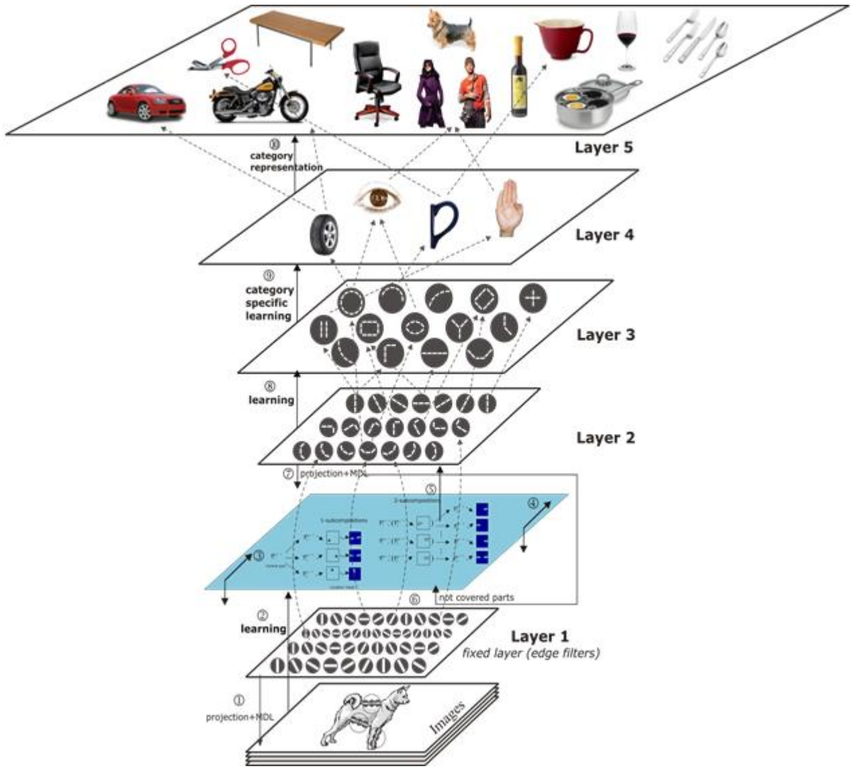


Latent space

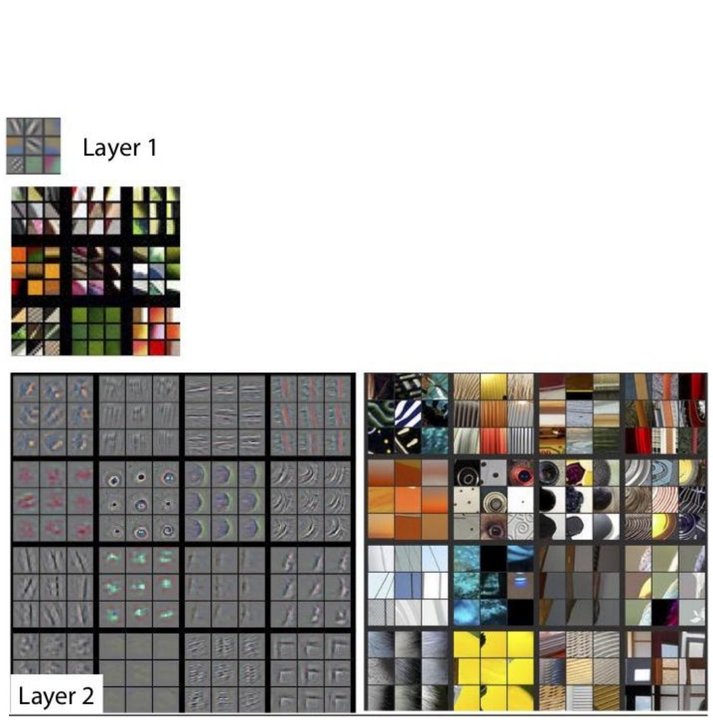
Deep *representations*

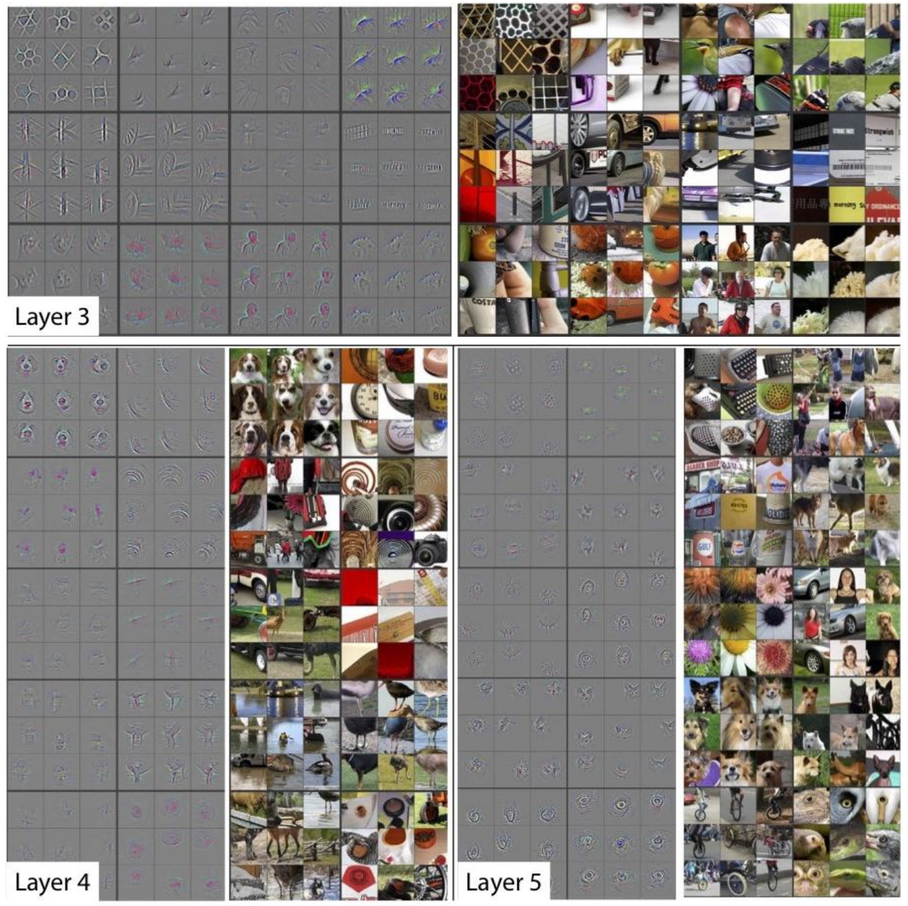
https://projector.tensorflow.org/

Deep *representations*



Deep *representations*



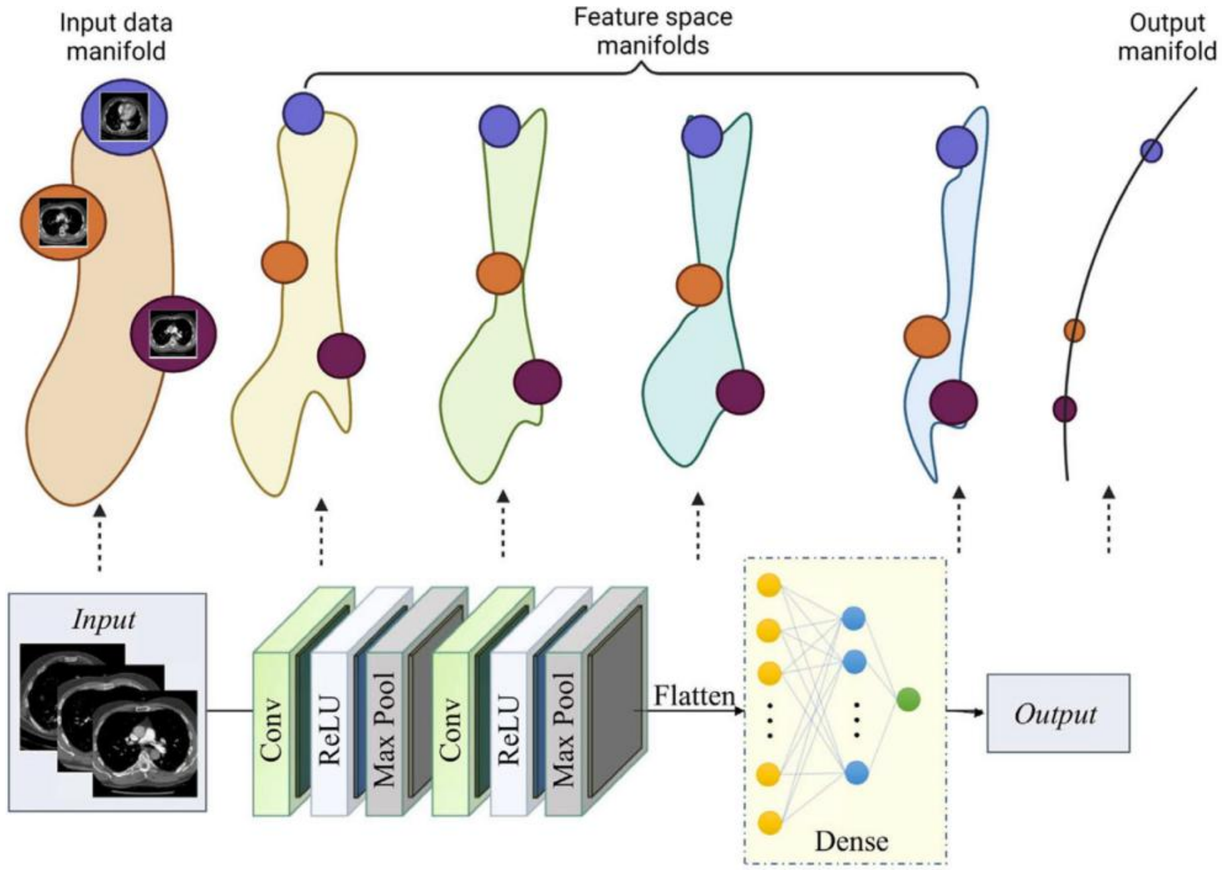


Deep *representations*

Edges (layer conv2d0) **Textures** (layer mixed3a) **Patterns** (layer mixed4a) **Parts** (layers mixed4b & mixed4c) **Objects** (layers mixed4d & mixed4e)

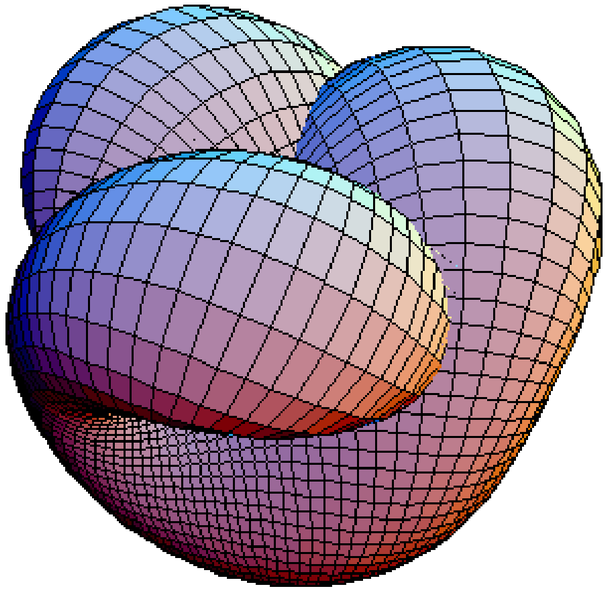
https://distill.pub/2017/feature-visualization/

Deep Neural Network *feature space*

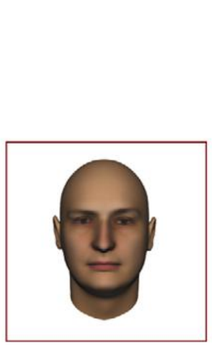


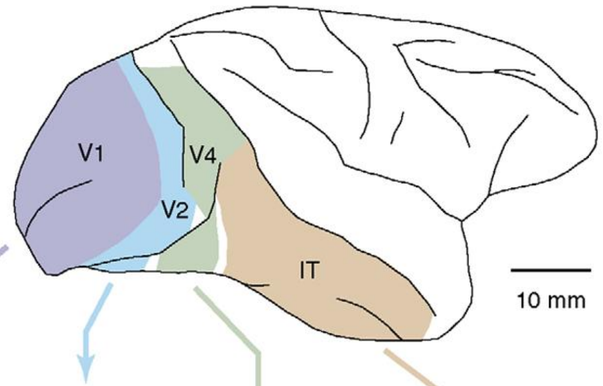
n-*manifold*

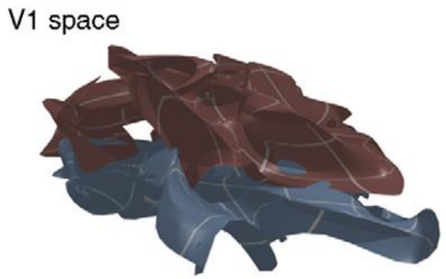
Un *manifold* es un **espacio Hausdorff segundo contable** que es **localmente homeomorfo** a un espacio euclídeo.

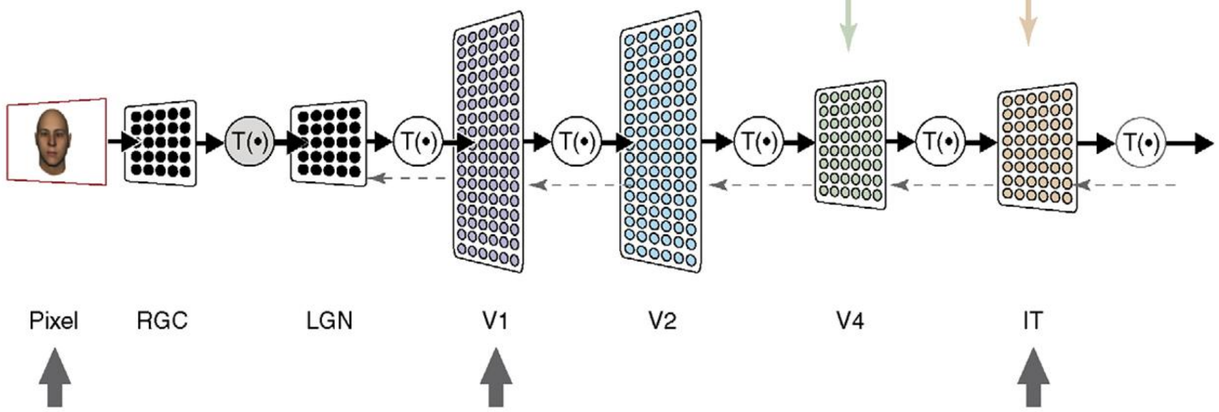


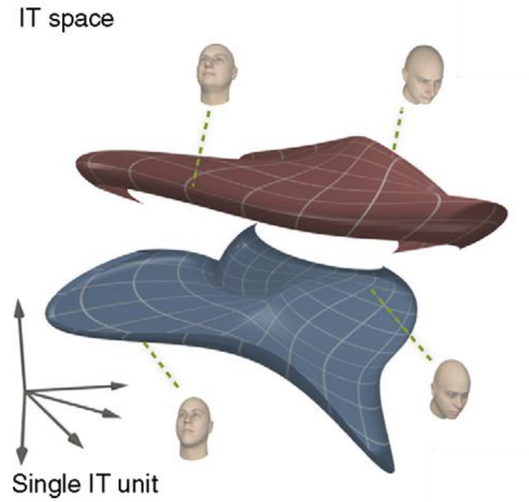
n-*manifold*



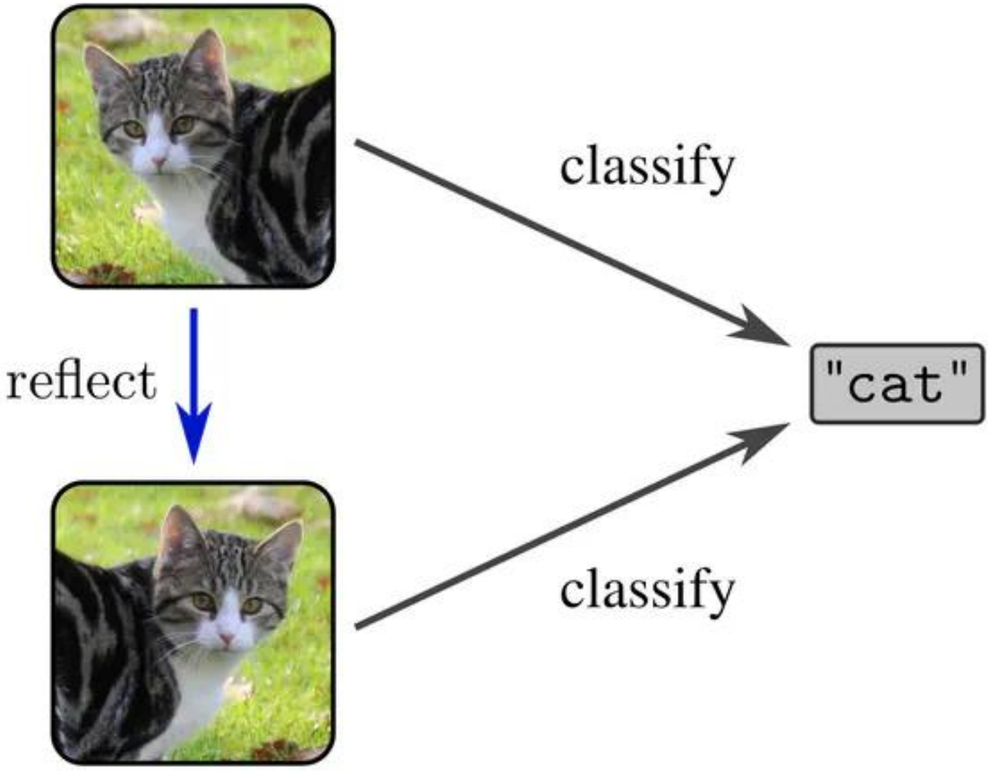

V1
V4
V2
IT
10 mm

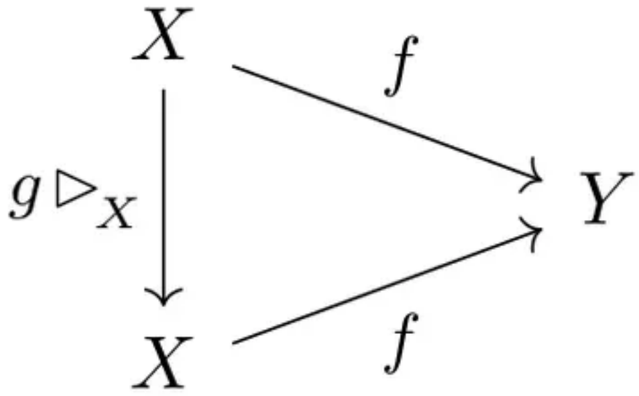

V1 space


T(•)
Pixel
RGC
LGN
V1
V2
V4
IT

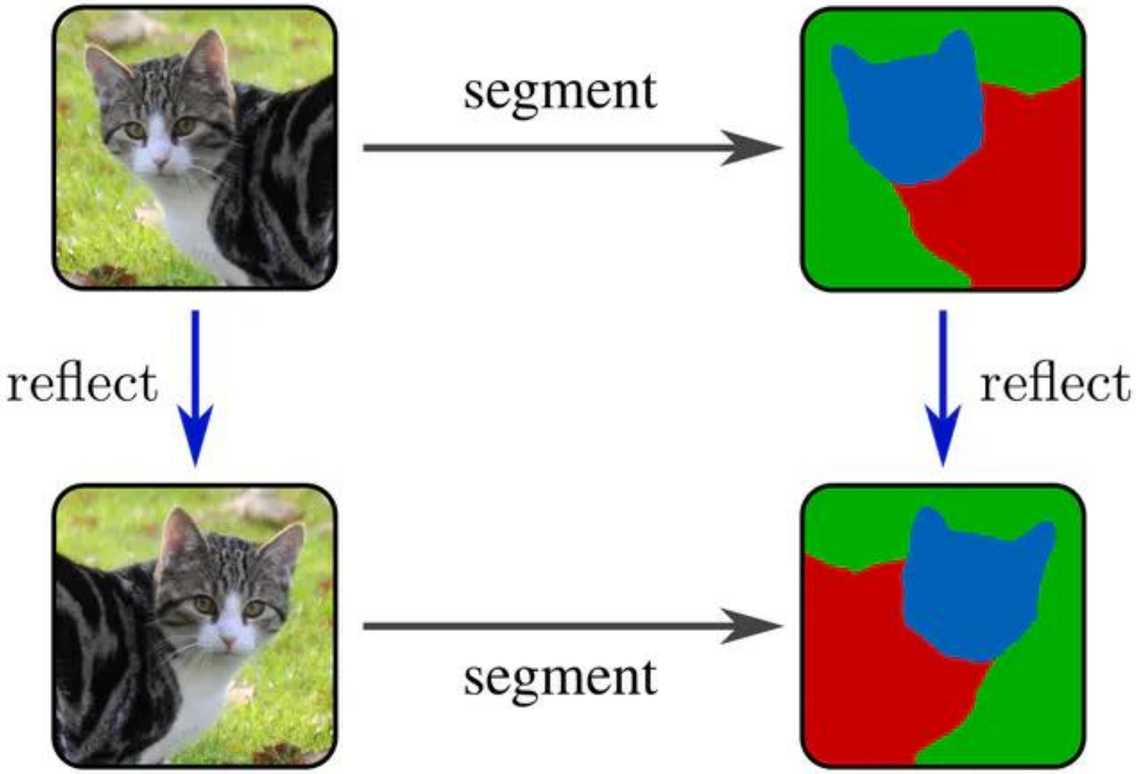

IT space
Single IT unit

Invariant *representation*



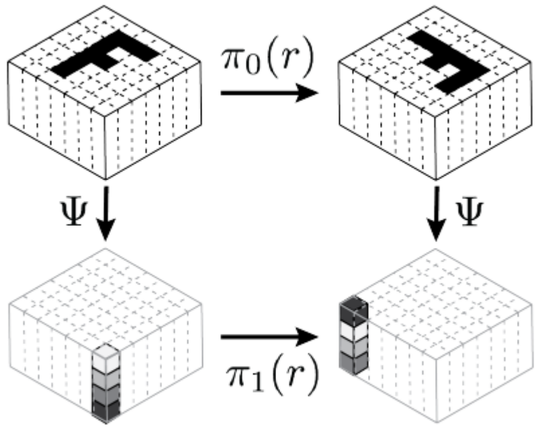


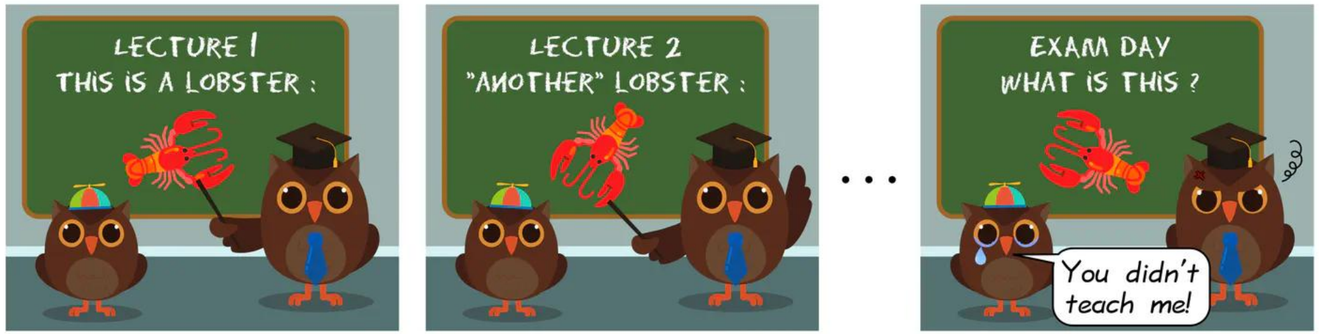
Equivariant *representation*

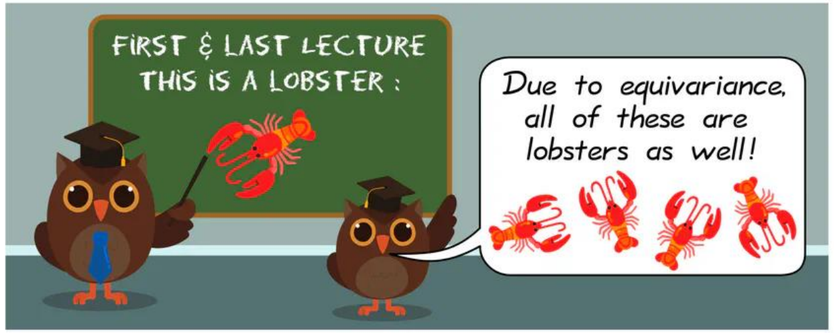


$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle g \triangleright_X}\downarrow & & \downarrow{\scriptstyle g \triangleright_Y} \\ X & \xrightarrow[f]{} & Y \end{array}$$

Equivariant *representation*

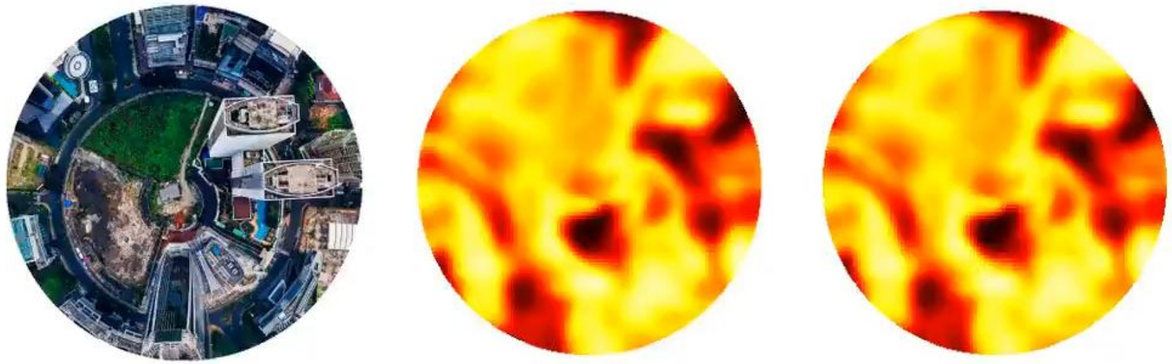




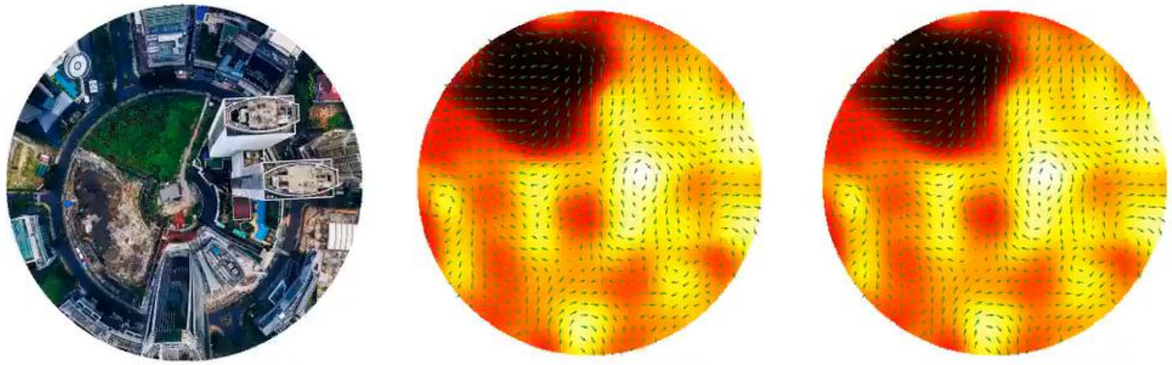


Equivariant *representation*

CNN



G-CNN



La convolución es equivariante de traslación, pero no de rotación.

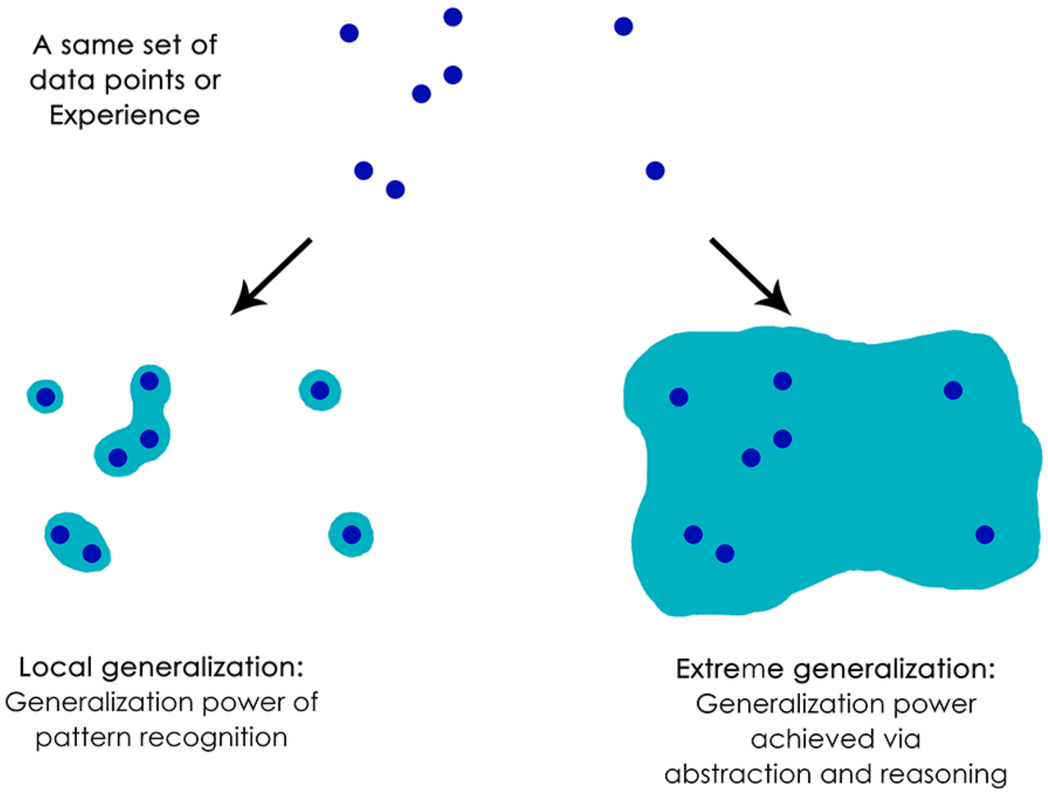
UTEC
UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

4.

Generalization

TRANSFORMATEC

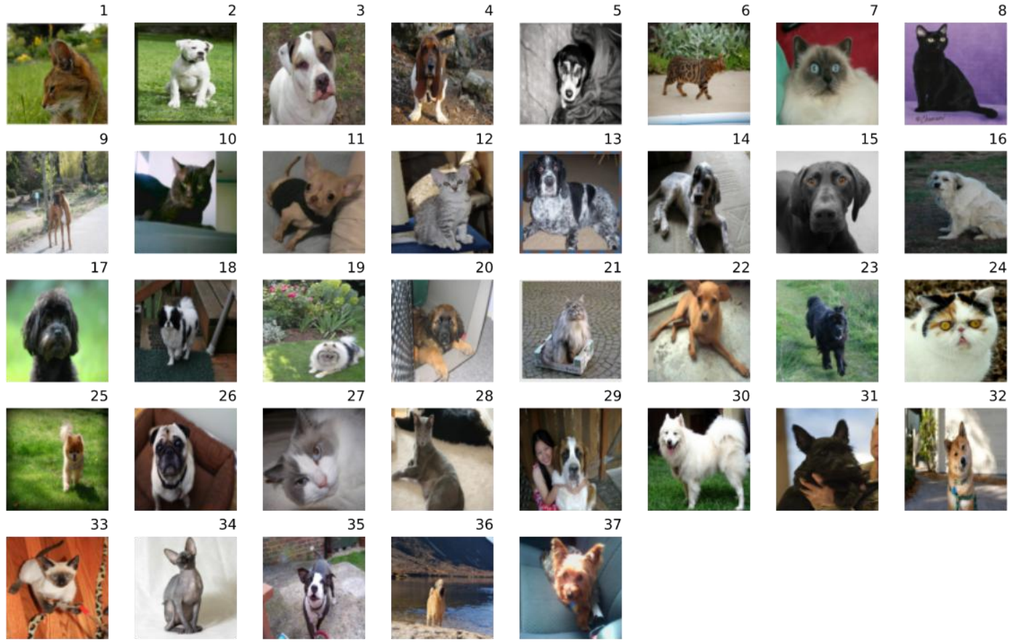
Generalization



In-Distribution *Generalization*

(IID Generalization)

El modelo es evaluado en datos que provienen de la misma distribución que los datos de entrenamiento, es decir, se asume que $P_{\text{train}(X,Y)} = P_{\text{test}(X,Y)}$.

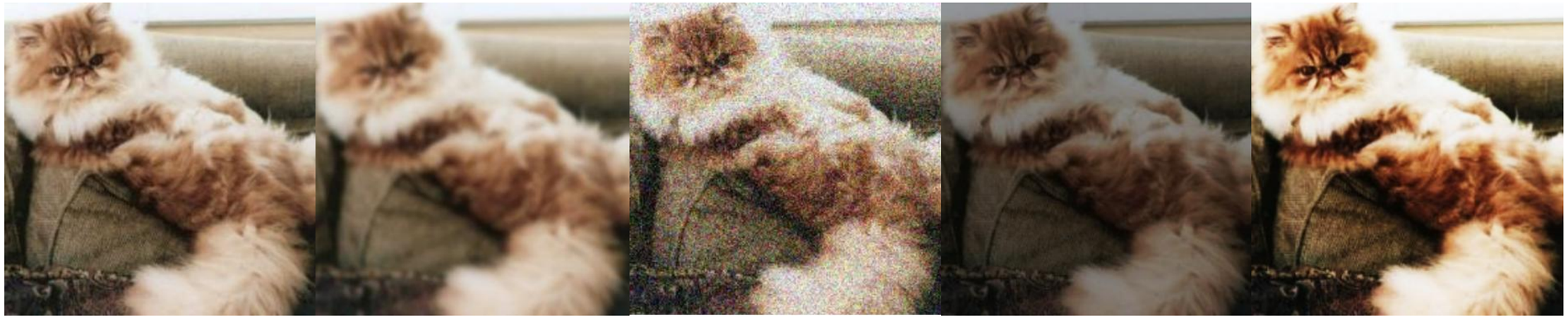


Training set

Test set

Intra-Domain *Generalization*

El modelo se evalúa dentro del mismo dominio, pero enfrentando variaciones que no estaban presentes durante el entrenamiento. Estas variaciones pueden incluir cambios en iluminación, orientación, ruido, etc.



Original **Blur** **Noise** **Brillo** **Contraste**

Out-of-Distribution *Generalization*

(OOD Generalization)

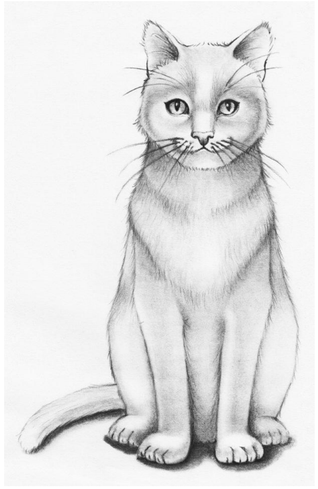
El modelo es evaluado con datos que provienen de una distribución diferente a la de entrenamiento, aunque dentro del mismo dominio conceptual






Training set



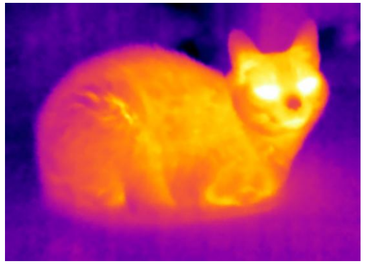


Test set

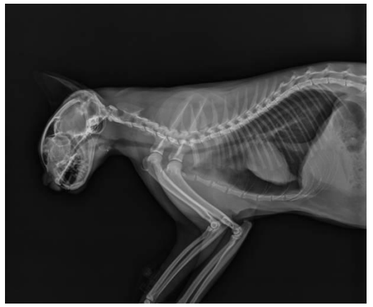
Cross-Domain *Generalization*

El modelo es entrenado en un dominio (p. ej., imágenes RGB) y evaluado en un dominio completamente diferente (p. ej., imágenes térmicas).





Training set

Test set

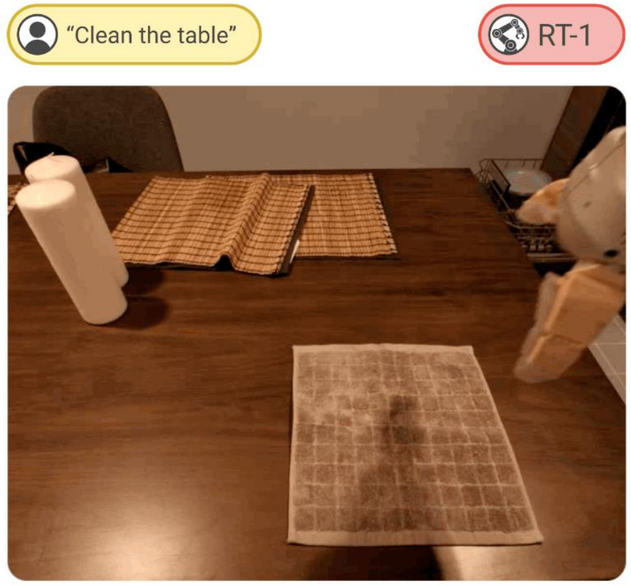
Open-Set Domain *Generalization*

En este caso, el modelo no solo debe enfrentarse a un nuevo dominio, sino que también debe reconocer clases nunca vistas durante el entrenamiento

Task *Generalization*

El modelo es entrenado para realizar una tarea específica y luego se evalúa en una tarea diferente pero relacionada, sin ser reentrenado

Training



Testing

Generalization

	Descripción	Diferencia clave	Ejemplo
In-Distribution	Misma distribución de entrenamiento	No hay cambios en la distribución	Entrenar y probar en imágenes de gatos similares
Intra-Domain	Mismo dominio con ligeras variaciones	Cambios menores dentro del dominio	Diferentes condiciones de iluminación
Out-of-Distribution (OOD)	Nueva distribución pero mismo dominio	Datos fuera de la distribución de entrenamiento	Dibujos, modelos 3D de gatos
Cross-Domain	Dominio completamente diferente	Cambio en la naturaleza de los datos	Imágenes térmicas de gatos
Open-Set Domain	Nuevo dominio + nuevas clases desconocidas	Nuevas clases en un entorno diferente	Clasificar nuevos animales nunca vistos
Task Generalization	Aprender en una tarea y generalizar a otra	Cambio de tarea	..

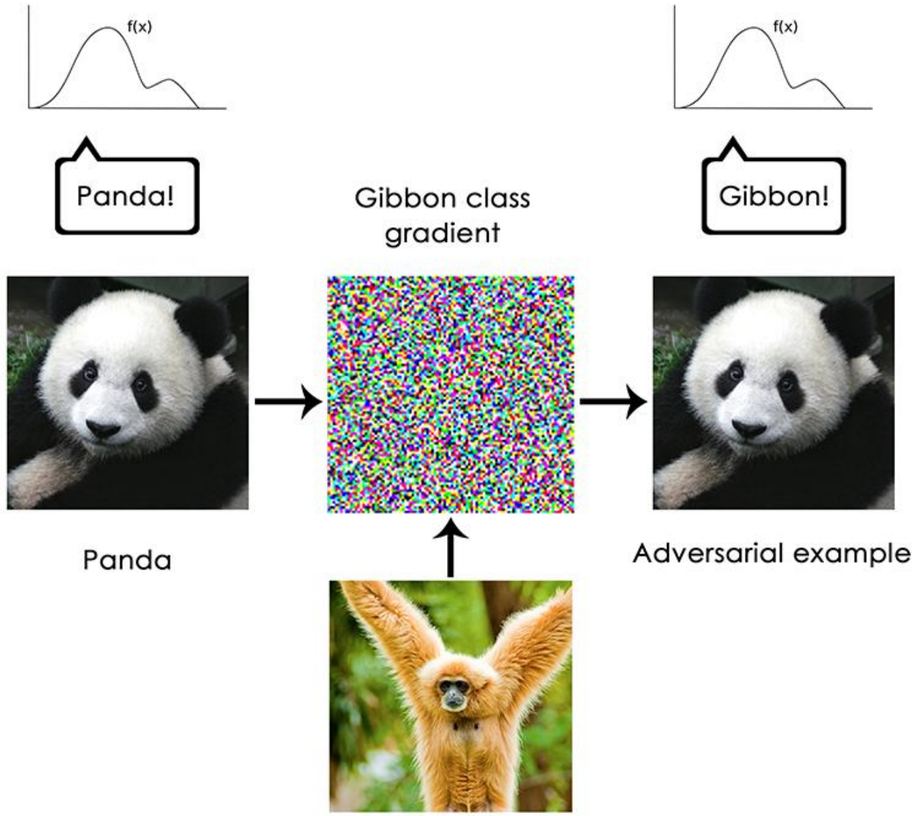
UTEC
UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

5.

Adversarial *Attack*

TRANSFORMATEC

Adversarial *Attack*



Adversarial *Attack*

White-box attacks

El atacante tiene acceso completo al modelo, incluyendo su arquitectura, parámetros y datos de entrenamiento. Esto permite generar ataques muy precisos y efectivos.

Fast Gradient Sign Method (FGSM)

$$x_{\text{adv}} = x + \epsilon \cdot \text{sign}\left(\nabla_x J(x, y)\right)$$

Donde:

- x entrada original (ej. imagen)
- y etiqueta verdadera de la entrada
- $J(x, y)$ función de pérdida del modelo
- $\nabla_x J(x, y)$ gradiente de la pérdida con respecto a la entrada x
- $\text{sign}(\cdot)$ función signo, que devuelve +1 o -1 por cada componente del gradiente
- ϵ parámetro de perturbación (determina la magnitud de la perturbación)

Ian J. Goodfellow et al. (2015) "Explaining and Harnessing Adversarial Examples" arXiv preprint arXiv:1412.6572.

Adversarial *Attack*

White-box attacks

El atacante tiene acceso completo al modelo, incluyendo su arquitectura, parámetros y datos de entrenamiento. Esto permite generar ataques muy precisos y efectivos.

Fast Gradient Sign Method (FGSM)

$$x_{\text{adv}} = x + \epsilon \cdot \text{sign}\big(\nabla_x J(x, y)\big)$$

Donde:

- x entrada original (ej. imagen)
- y etiqueta verdadera de la entrada
- $J(x, y)$ función de pérdida del modelo
- $\nabla_x J(x, y)$ gradiente de la pérdida con respecto a la entrada x
- $\text{sign}(\cdot)$ función signo, que devuelve +1 o -1 por cada componente del gradiente
- ϵ parámetro de perturbación (determina la magnitud de la perturbación)

Projected Gradient Descent (PGD)

$$x_{t+1} = Proj_{B_\epsilon(x)}\left(x_t + \alpha \cdot sign\big(\nabla_x J(x, y)\big)\right)$$

Donde:

- x_t entrada adversarial en la iteración actual
- α paso de actualización en cada iteración (usualmente $\alpha < \epsilon$)
- $\text{sign}\big(\nabla_x J(x, y)\big)$ dirección del gradiente de la pérdida respecto a la entrada
- $\text{Proj}_{B_\epsilon(x)}$ proyección en una ball L_∞ de radio ϵ centrada en x, asegurando que la perturbación no supere el límite establecido

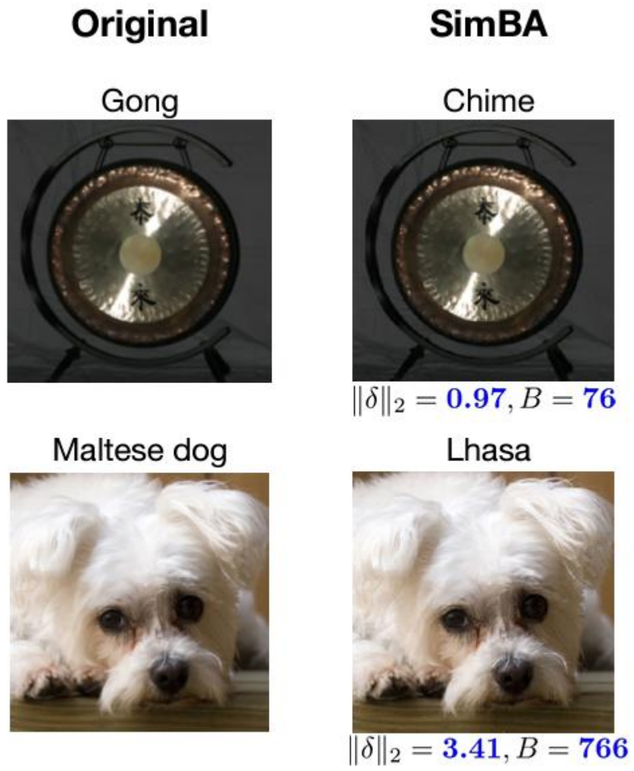
Adversarial *Attack*

Black-box attacks

El atacante no tiene acceso al modelo, solo puede observar las respuestas del modelo a diferentes entradas.

Algorithm 1 SimBA in Pseudocode

procedure SIMBA($\mathbf{x}, y, Q, \epsilon$)
  $\delta = \mathbf{0}$
  $\mathbf{p} = p_h(y \mid \mathbf{x})$
  while $\mathbf{p}_y = \max_{y'} \mathbf{p}_{y'}$ **do**
    Pick randomly without replacement: $\mathbf{q} \in Q$
    for $\alpha \in \{\epsilon, -\epsilon\}$ **do**
      $\mathbf{p}' = p_h(y \mid \mathbf{x} + \delta + \alpha\mathbf{q})$
      if $\mathbf{p}'_y < \mathbf{p}_y$ **then**
        $\delta = \delta + \alpha\mathbf{q}$
        $\mathbf{p} = \mathbf{p}'$
        break
  return δ



Chuan Guo et al. (2019) "Simple Black-box Adversarial Attacks" International conference on machine learning (pp. 2484-2493). PMLR.

UTEC
UNIVERSIDAD DE INGENIERÍA Y TECNOLOGÍA

> Reinventa el mundo <

GRACIAS

Victor Flores Benites

TRANSFORMATEC