Variational AutoEncoders invariantes frente a cambios de distribución

Matias Vera [1,2] Martin Gonzalez [1] Leonardo Rey Vega [1,2]

[1]Facultad de Ingeniería, Universidad de Buenos Aires [2]Centro de Simulación Computacional para Aplicaciones Tecnológicas, CONICET

Invarianza

Los algoritmos de aprendizaje tienden a absorber imprudentemente todas las correlaciones encontradas en los datos de entrenamiento.

(A) **Cow: 0.99**, Pasture: 0.99, Grass: 0.99, No Person: 0.98, Mammal: 0.98

(B) No Person: 0.99, Water: 0.98, Beach: 0.97, Outdoors: 0.97, Seashore: 0.97

(C) No Person: 0.97, **Mammal: 0.96**, Water: 0.94, Beach: 0.94, Two: 0.94

Figure 1. Ejemplo de aprendizaje espúreo [Beery et al. 2018].

La línea de trabajo a seguir en este plan tiene que ver con el desarrollo de algoritmos que aprendan solamente características invariantes a cambios del entorno. Para ellos se contará con algunos entornos de entrenamiento diferentes entre sí y diferentes al entorno de testeo. El objetivo es aprender solamente las características presentes en todos los entornos.

Variational Autoencoders

Los VAE [Kingma and Welling 2014] poseen dos objetivos: 1) Reconstruir los datos a partir de una representación relevante, y 2) Permitir la generación de datos sintéticos distribuidos como los datos de entrenamiento. Para ello utilizan el enfoque bayesiano de suponer la cadena de Markov $X \to Z \to \hat{X}$ con

- $Z|_{X=x} \sim \mathcal{N}(\mu_\theta(x), \sigma^2_\theta(x))$ (encoder)
- $\hat{X}|_{Z=z} \sim$ Ber$(f_\theta(z))$ o $\hat{X}|_{Z=z} \sim \mathcal{N}(f_\theta(z), I)$ (decoder)
- $\tilde{Z} \sim \mathcal{N}(0, I)$ (prior)

donde $\mu_\theta, \sigma^2_\theta$ y f_θ suelen ser redes neuronales. Para entrenar se minimiza una función costo [Higgins et al. 2017] que combine sus dos objetivos:

$$\underbrace{\mathbb{E}[-\log p_{\hat{X}|Z}(x|Z)|X=x]}_{\text{Reconstrucción}} + \beta\cdot\underbrace{\mathcal{D}(p_{Z|X}(\cdot|x)\|p_{\tilde{Z}})}_{\text{Generación}}$$

Ejemplo 1: Colored MNIST

Se colorea de rojo o verde los dígitos de MNIST con diferentes probabilidades definiendo dos entornos de entrenamiento y uno de testeo [Arjovsky et al. 2019]:

	$\mathcal{E}^{(1)}_{\text{train}}$	$\mathcal{E}^{(2)}_{\text{train}}$	$\mathcal{E}^{(1)}_{\text{test}}$
$\mathbf{P}(\text{"verde"}\,\|\,y \leq 4)$	0.7	0.65	0.3
$\mathbf{P}(\text{"verde"}\,\|\,y \geq 5)$	0.3	0.35	0.7

Selección de métricas para evaluación

Es interesante elegir unas métricas objetivas para analizar el desempeño de un algoritmo independientes del modelo utilizado. Es por eso que se eligieron:

1. **MSE para reconstrucción:** Dado que se entrenó con unidades Bernoulli, utilizar como métrica de reconstrucción el error cuadrático medio parece la opción indicada.
2. **RKL para generación:** Para la generación se opto por la métrica Reverse Kullback Leibler. La decisión de esto es porque se está comparando una generación *invariante* contra un conjunto de testeo con un sesgo diametralmente opuesto a los conjuntos de entrenamiento, y la RKL posee un sesgo hacia generación exclusiva.
3. **Diferencia de MSE para invarianza:** Dadas las proporciones de coloreo en los conjuntos de entrenamiento, será más fácil reconstruir números verdes pequeños y números rojos grandes que números verdes grandes y números rojos pequeños. Es por eso que se decidió por computar la diferencia relativa entre los errores cuadráticos medios de los datos *difíciles* y los *fáciles* sobre el error *fácil*.

Resultados en Colored MNIST

Métrica	Múltiple Decoders	Múltiple Encoders
MSE	14.7	12.8
RKL	1.957	2.839
$\mathcal{E}_{\text{rel}}$	0.100	0.121

Invariant Variational Autoencoder: Múltiples decoders

A los objetivos de reconstrucción y generación se le suma el objetivo de la invarianza, definiendo el modelo como

- $Z|_{X=x} \sim \mathcal{N}(\mu_\theta(x), \sigma^2_\theta(x))$ *(encoder)*
- $\hat{X}|_{Z=z} \sim$ *Ber*$(f_\theta(z))$ *o* $\hat{X}|_{Z=z} \sim \mathcal{N}(f_\theta(z), I)$ *(decoder invariante)*
- $\hat{X}|_{\substack{Z=z\\E=e}} \sim$ *Ber*$(g_\theta(z,e))$ *o* $\hat{X}|_{\substack{Z=z\\E=e}} \sim \mathcal{N}(g_\theta(z,e), I)$ *(decoders variantes)*
- $\tilde{Z} \sim \mathcal{N}(0, I)$ *(prior)*

Para entrenar se minimiza una función costo que combine sus tres objetivos:

$$\underbrace{\mathbb{E}[-\log p_{\hat{X}|Z,E}(x|Z,e)|X=x]}_{\text{Reconstrucción}} + \beta\cdot\underbrace{\mathcal{D}(p_{Z|X}(\cdot|x)\|p_{\tilde{Z}})}_{\text{Generación}} + \gamma\cdot\underbrace{\mathbb{E}[-\log p_{\hat{X}|Z}(x|Z)|X=x]}_{\text{Invarianza}}$$

Invariant Variational Autoencoder: Múltiples Encoders

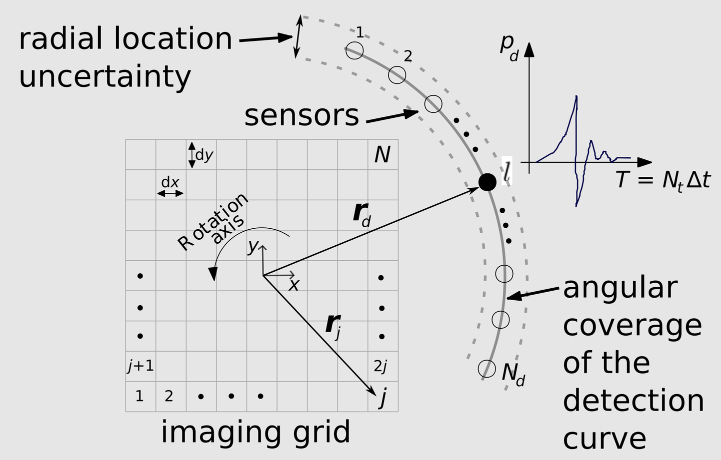
A los objetivos de reconstrucción y generación se le suma el objetivo de la invarianza, definiendo el modelo como

- $Z|_{X=x} \sim \mathcal{N}(\mu_\theta(x), \sigma^2_\theta(x))$ (encoder invariante)
- $Z|_{\substack{Z=z\\E=e}} \sim \mathcal{N}(\mu_\theta(x), s^2_\theta(x,e))$ (encoders variantes)
- $\hat{X}|_{Z=z} \sim$ Ber$(f_\theta(z))$ o $\hat{X}|_{Z=z} \sim \mathcal{N}(f_\theta(z), I)$ (decoder)
- $\tilde{Z} \sim \mathcal{N}(0, I)$ (prior)

Para entrenar se minimiza una función costo que combine sus tres objetivos (muestreando con los encoder variantes):

$$\underbrace{\mathbb{E}[-\log p_{\hat{X}|Z}(x|Z)|X=x]}_{\text{Reconstrucción}} + \beta\cdot\underbrace{\mathcal{D}(p_{Z|X}(\cdot|x)\|p_{\tilde{Z}})}_{\text{Generación}} + \gamma\cdot\underbrace{\mathcal{D}\left(p_{Z|X,E}(\cdot|x,e)\|p_{Z|X}(\cdot|x)\right)}_{\text{Invarianza}}$$

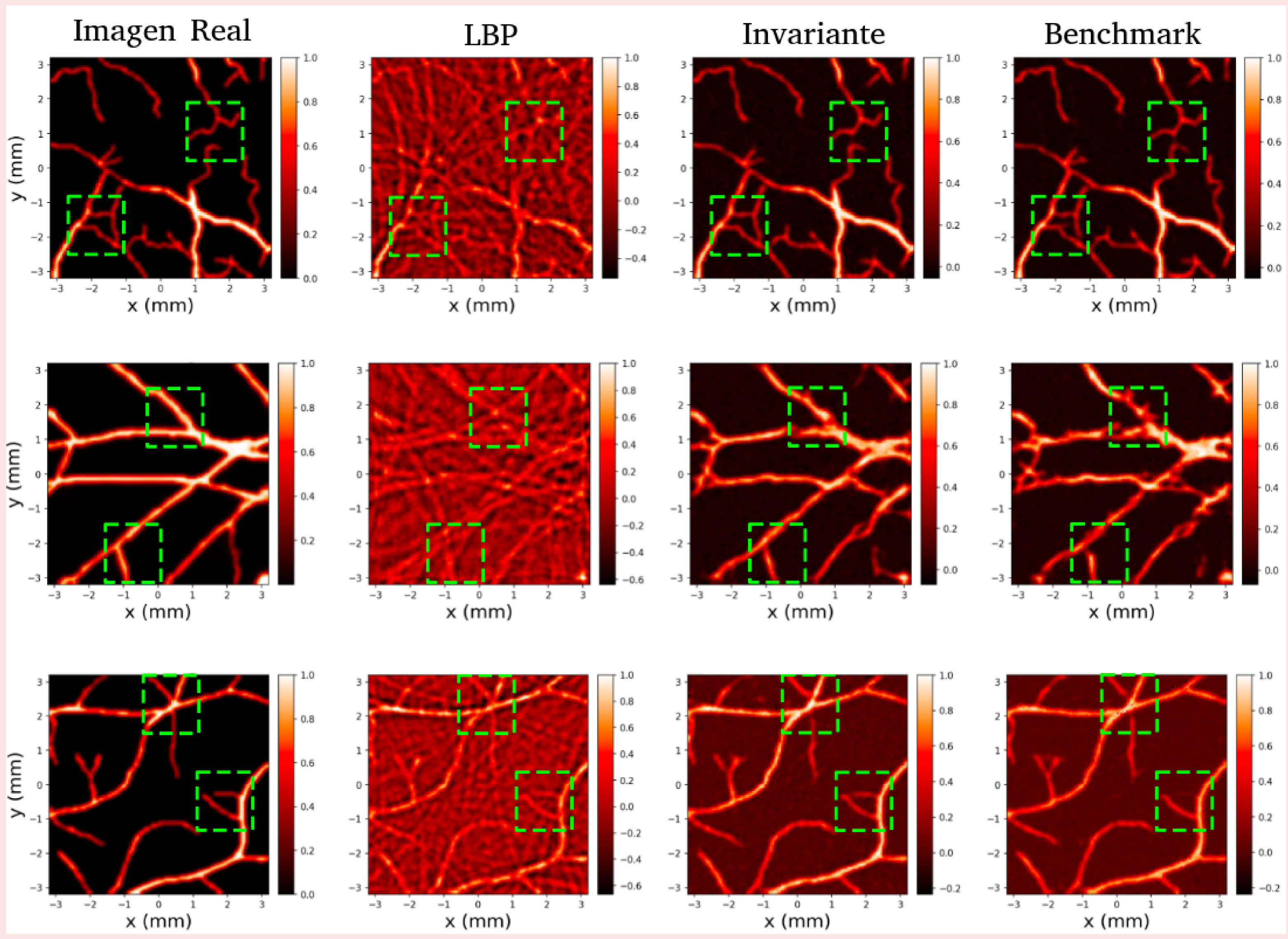
Ejemplo 2: Tomografía Optoacústica



Para independizarse del setup experimental, se construyo una versión *denoising* de un autoencoder invariante capaz de mejorar la resolución de la imagen sensada.

Resultados en Tomografía Optoacústica

Un estudio cualitativo registró ciertas ventajas de utilizar este enfoque para diferentes incertezas en las posición de los sensores. Para más información consultar [Vera et al. 2023].



References

- Vera, M., M. G. González, and L. Rey Vega (2023). "Invariant representations in deep learning for optoacoustic imaging". In: *Review of Scientific Instruments.*
- Arjovsky, M., L. Bottou, I. Gulrajani, and D. Lopez-Paz (2019). "Invariant Risk Minimization". In: *ArXiv* abs/1907.02893.
- Beery, S., G. Van Horn, and P. Perona (2018). "Recognition in Terra Incognita". In: *European Conference on Computer Vision (ECCV).*
- Higgins, I., L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner (2017). "β-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework". In: *International Conference on Learning Representations (ICLR).*
- Kingma, D. P. and M. Welling (2014). "Auto-Encoding Variational Bayes". In: *International Conference on Learning Representations (ICLR).*