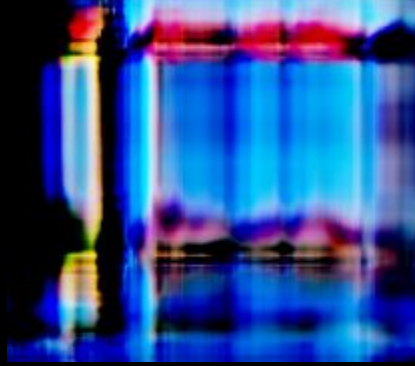
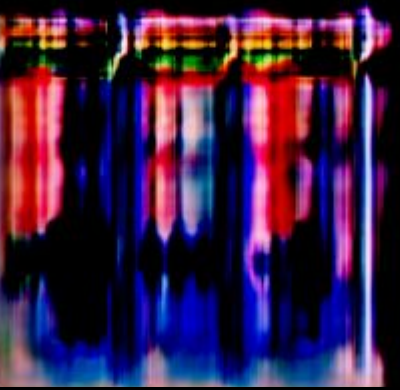
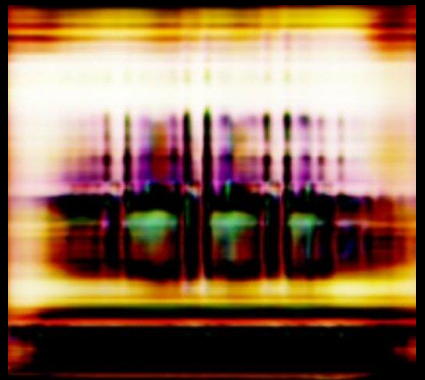
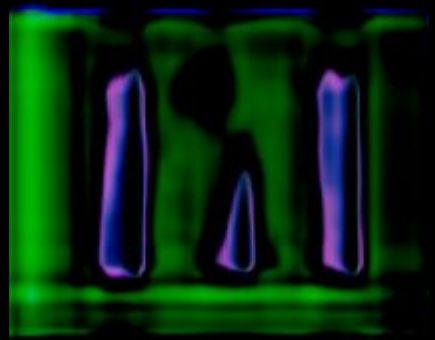


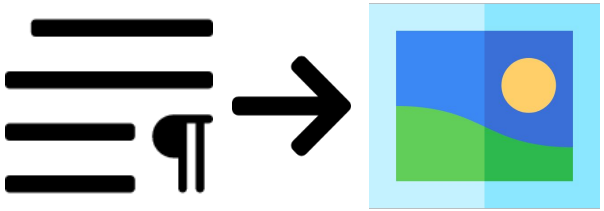



Text-based image generation

Paola Mejía, Elizabeth Rodríguez, Juan B. Martínez Parente


Problema a resolver



Generar imágenes realistas a partir de oraciones que las describen.

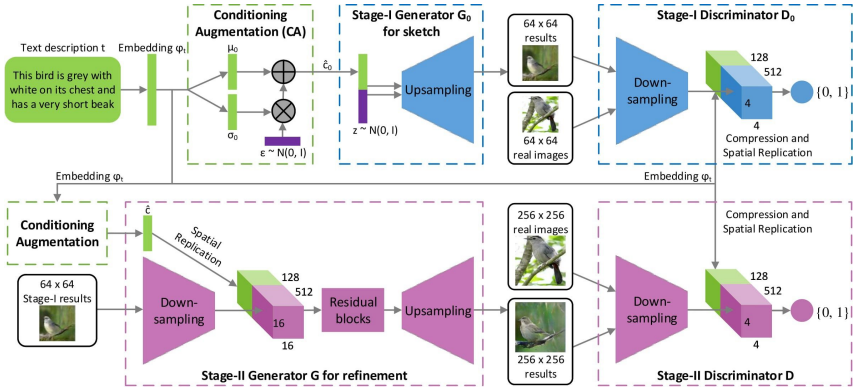
A group of teenage boys on a road jumping joyfully.
Four boys running and jumping
Four kids jumping on the street with a blue car in the back
Four young men are running on a street and jumping for joy
Several young men jumping down the street

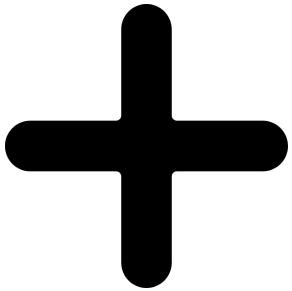
Los datos

Flickr 8K

- **8,092** imagenes
- principalmente de personas y perros realizando diversas actividades
- Cada una con **5** anotaciones que la describe

Solución propuesta

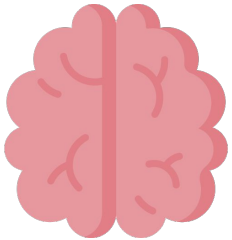




- Optimizador Adam
- Tasa de aprendizaje de 0.0002
- Función de activación: LeakyReLU
- Tamaño de *batch* de 128

Representación latente

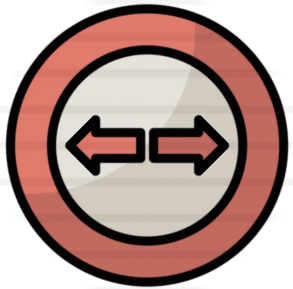
Bidirectional Encoder Representations from Transformers



Modelo de lenguaje natural pre-entrenado

Desarrollado por Google en 2018

Genera una representación de cada palabra a partir de otras palabras en la frase



Bidireccional: considera el contexto tanto a la derecha como a la izquierda de la palabra

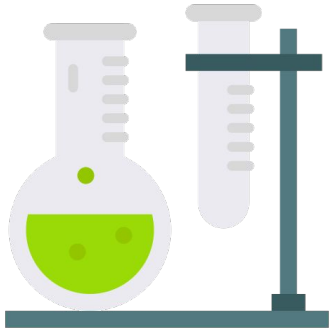
Tabla 2: Frase de prueba: "A man is cycling on the road"

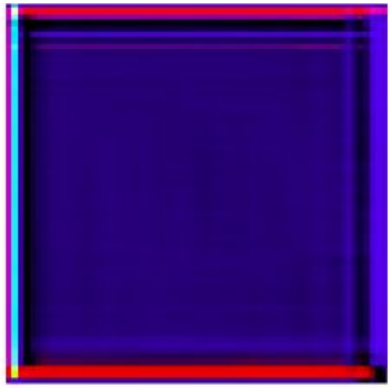
Frases del entrenamiento	Distancia coseno
A man riding his bike in traffic	0.057
A man riding a bike near traffic	0.062
A man holds an object with his hand while riding his bike down the street	0.074
Man riding a bicycle down a road with clouds overhead	0.075

Tabla 3: Frase de prueba: "A child painting a picture"

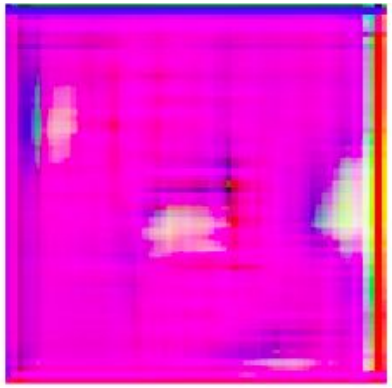
Frases del entrenamiento	Distancia coseno
A child painting on a piece of spinning paper	0.092
A young girl painting a picture	0.110
A child paints with different colors using brushes	0.11
A child plays with paint	0.151

Experimentos

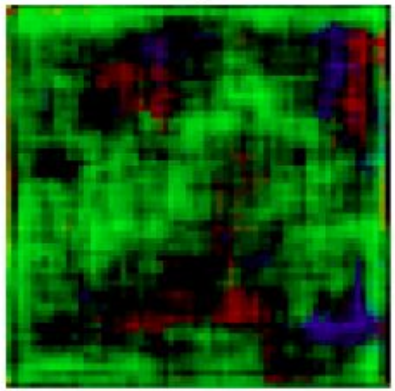




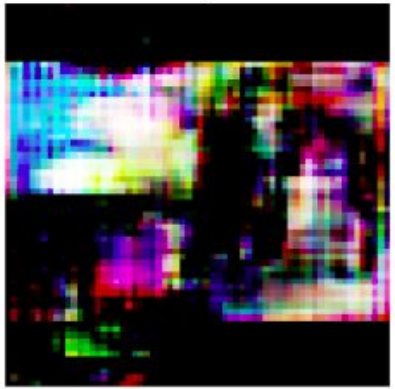
(a) Tokenizer simple



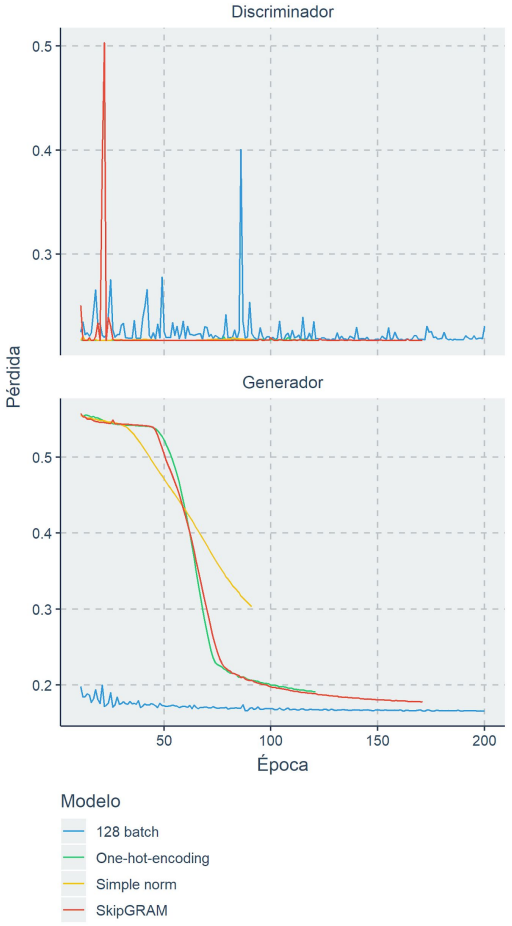
(b) One hot encoding

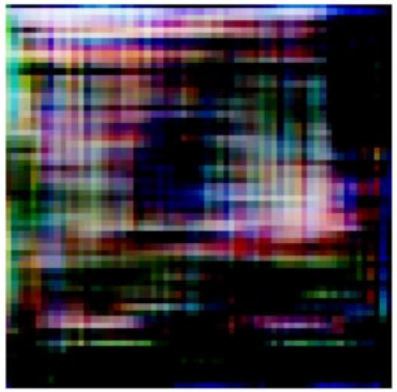


(c) Skipgram

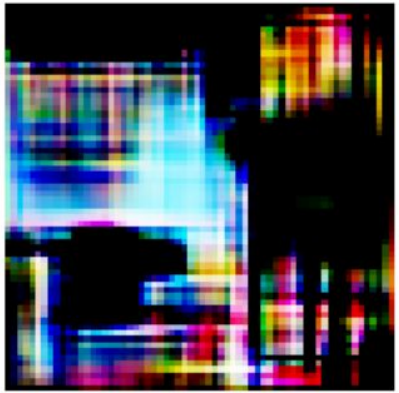


(d) Bert



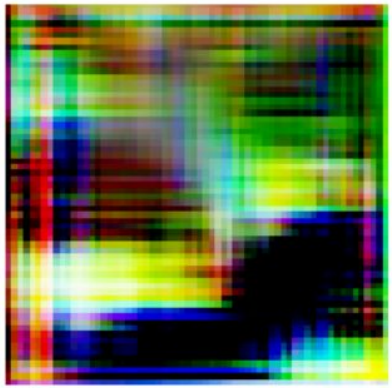


(a) 32

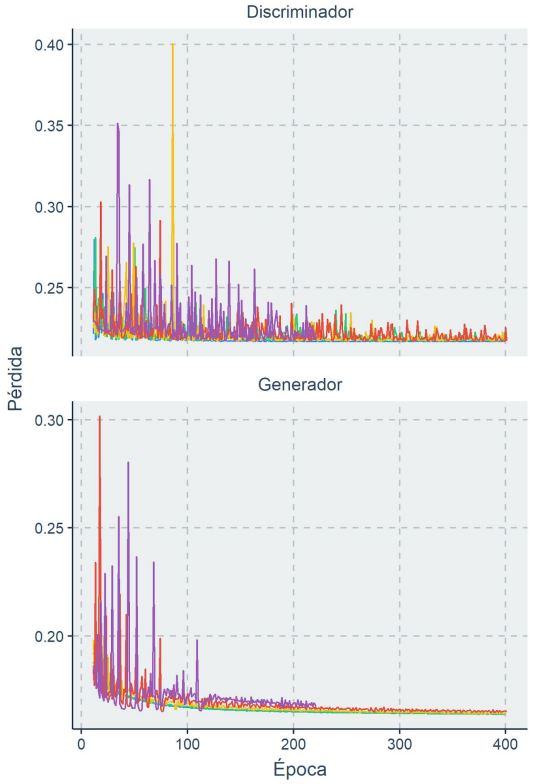


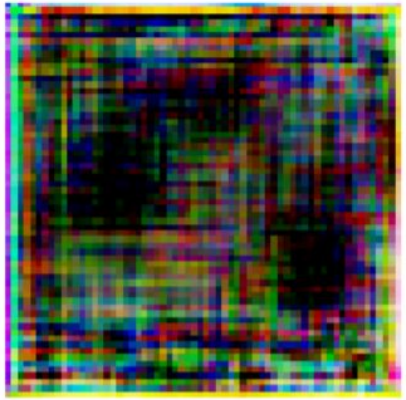
(b) 64

(c) 128

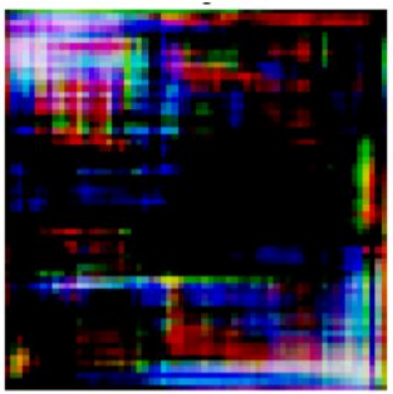


(d) 512

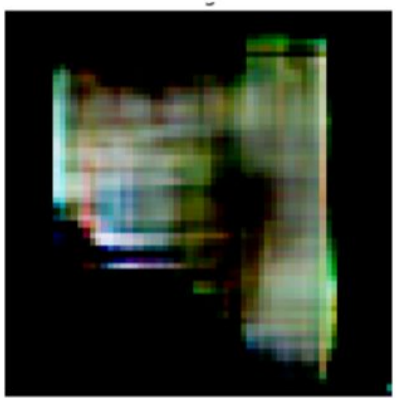




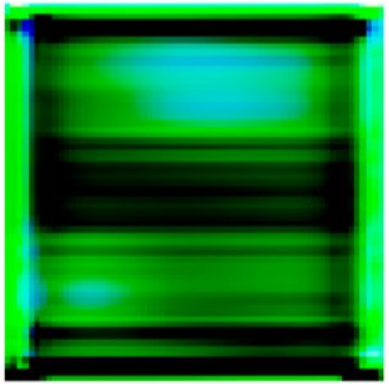
(a) Mas chico: 0.0002



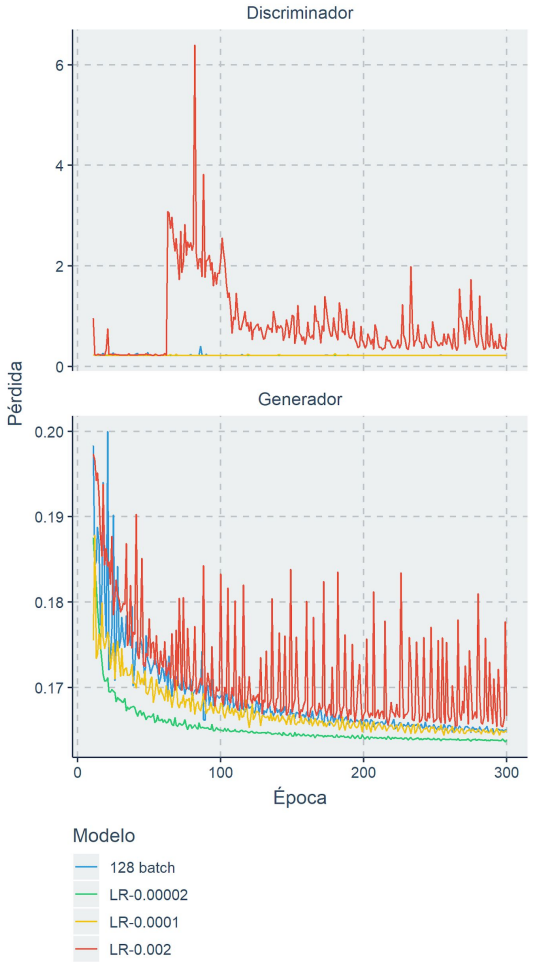
(b) 0.0001

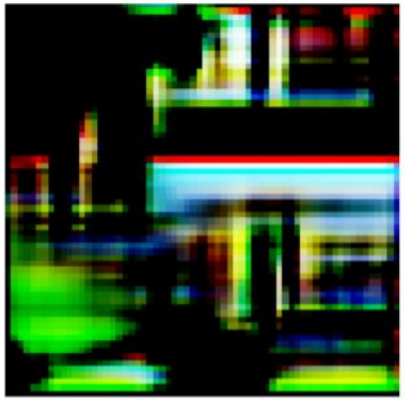


(c) 0.0002

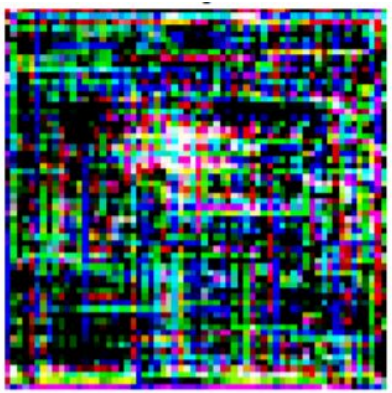


(d) Mas grande: 0.002

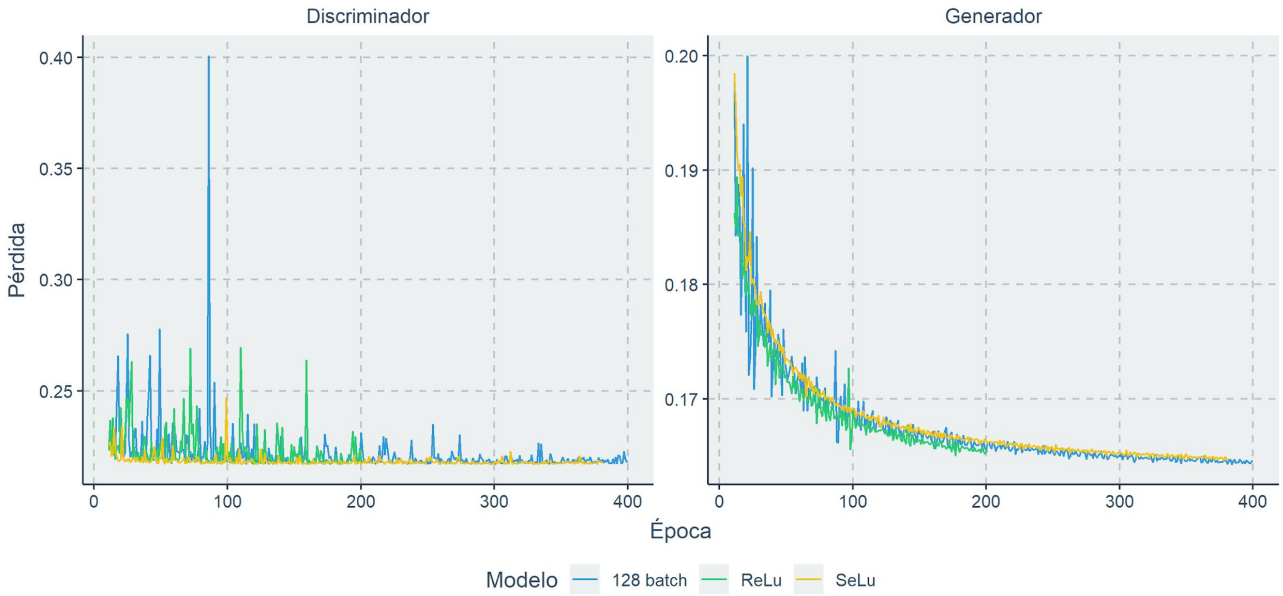




(a) ReLU

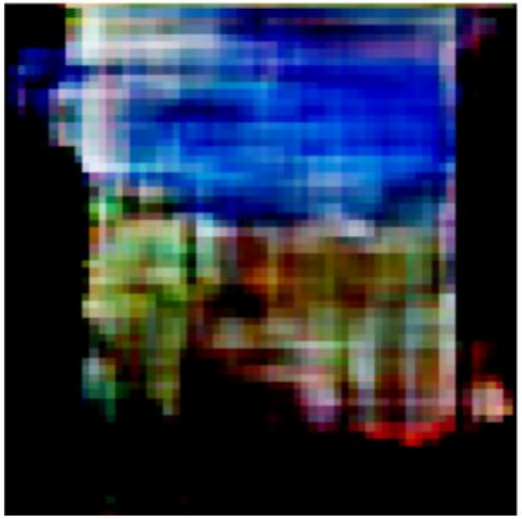


(b) SeLU

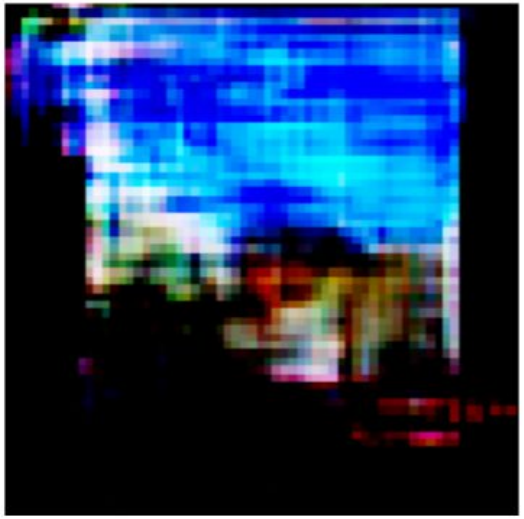

Discriminador
Generador
Pérdida
0.40
0.35
0.30
0.25
0.20
0.19
0.18
0.17
0
100
200
300
400
Época
Modelo
128 batch
ReLu
SeLu

Resultados

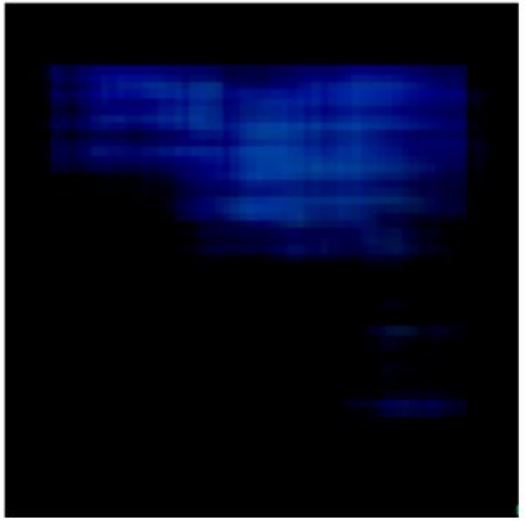
Resultados (Fase I)



(a) bird descend **sky**

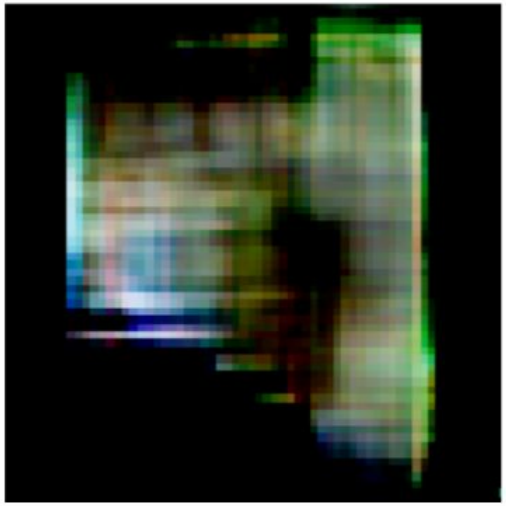


(b) three person wear parachute free falling together **sky**

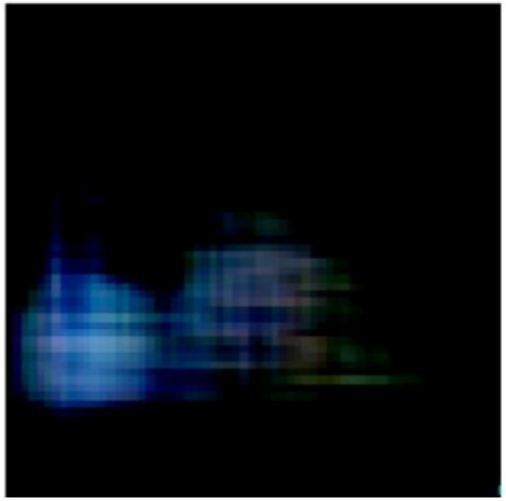


(c) several person stand top snowy rocky hill moon **sky**

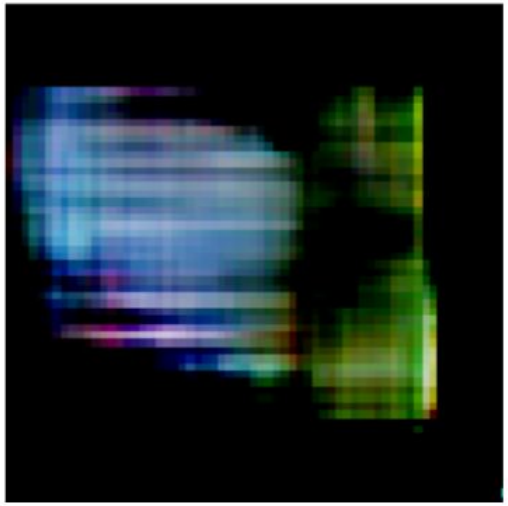
Resultados (Fase I)



(a) boy black hair dark skin swim murky **water**

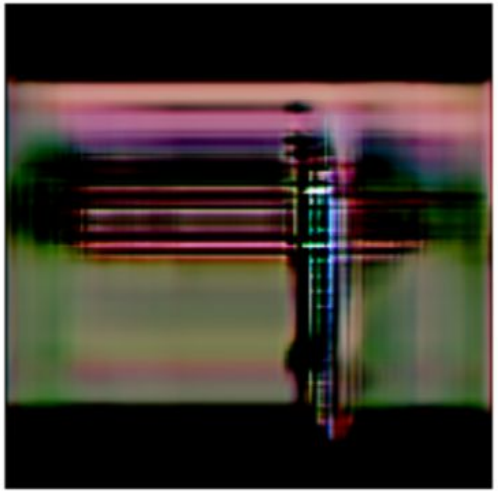


(b) man flip white **water** raft

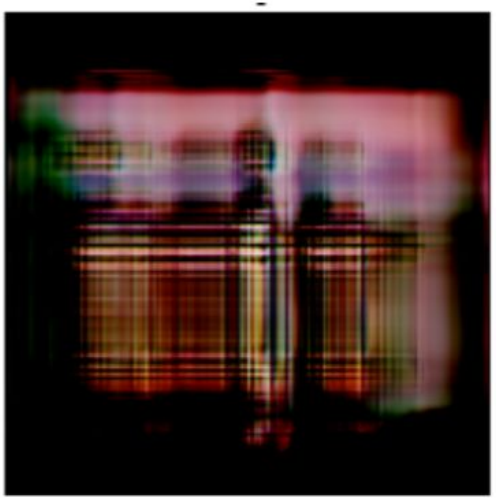


(c) man tag line go **water**

Resultados (Fase II)

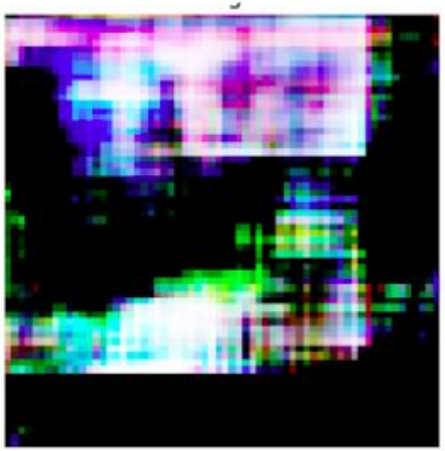


(a) black dog jumps into the water holding stick in the mouth

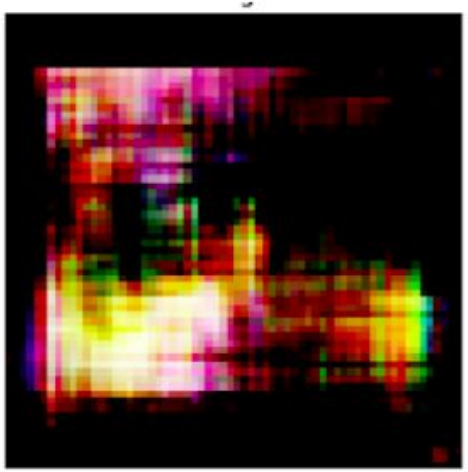


(b) little girl playing in a playground

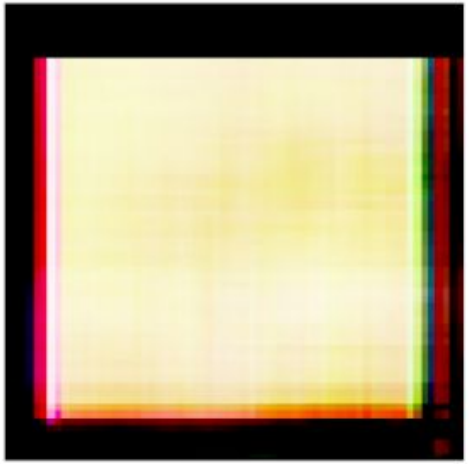
Aleatorios



(a) white clouds above river

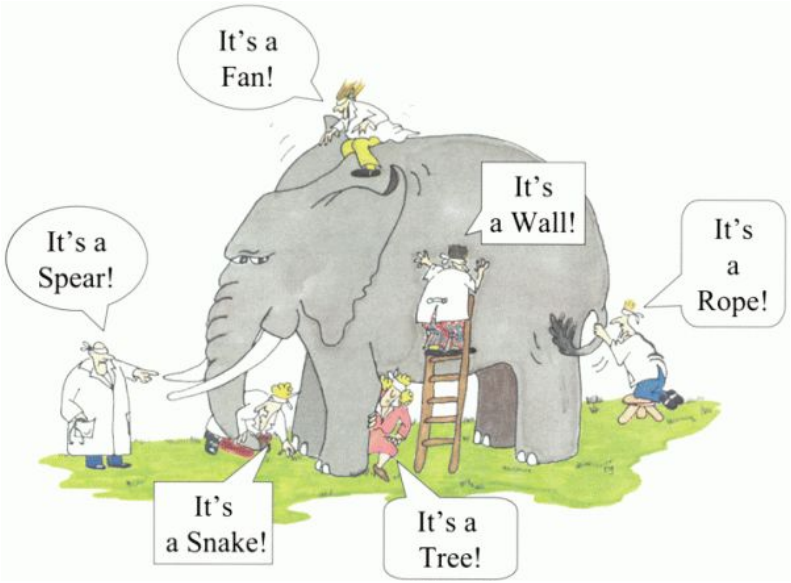


(b) red



(c) Juan, Ely and Paola finishing the project

Conclusiones

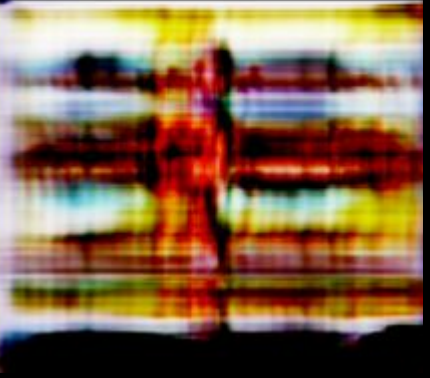


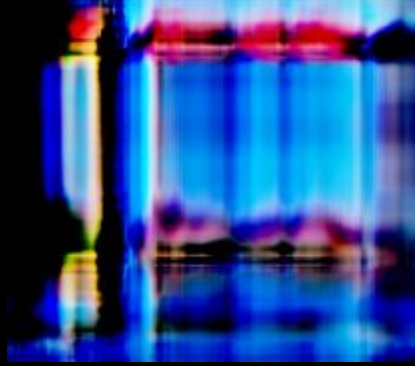
Limitaciones del modelo

No se logran distinguir con claridad las formas concretas, se ven sombras y manchas.

Aprendizajes

Multiplicidad de soluciones existentes (y muchas más por descubrir)



Preguntas