
Memoria de recreación
de estilos artísticos



UNIVERSIDAD
COMPLUTENSE
MADRID

UNIVERSIDAD COMPLUTENSE DE MADRID

Facultad de Ciencias Matemáticas

Asignatura: Taller de tecnomatemática

Profesor: Antonio López Montes

Autores: Mario Bermejo Cuervo, Vanessa Carnero Olivares, Jesús Gutiérrez Díaz
y Aroa Ortiz Lominchar

Abril de 2026

Capítulo 1

Introducción al proyecto

1.1. Planteamiento

Este proyecto toma como punto de partida el trabajo *LouvreSAE: Sparse Autoencoders for Interpretable and Controllable Style Transfer* [2] y lo traslada a una implementación utilizable desde un notebook. La idea general es abordar la transferencia de estilo artístico sin volver a entrenar un modelo distinto para cada pintor o corriente, sino reutilizando una arquitectura ya preparada que permite extraer un perfil de estilo a partir de varias obras de referencia y aplicarlo después a una imagen diferente. En ese sentido, nuestro trabajo no consiste en rediseñar el método, sino en convertirlo en un flujo de uso claro, directo y práctico.

1.2. Relación con el paper

La aportación central del paper consiste en representar el estilo como un conjunto de conceptos latentes que aparecen de manera consistente en varias obras de un mismo autor o de una misma familia visual [2]. Esa formulación evita tratar el estilo como una etiqueta abstracta o como un simple texto de entrada, y lo convierte en una estructura manipulable dentro del espacio latente del modelo. El notebook `tecno-louvre.ipynb` adopta esa misma filosofía: parte del sistema ya construido por los autores, descarga sus componentes y ejecuta la parte de inferencia necesaria para producir una imagen estilizada a partir de entradas elegidas por el usuario.

1.3. Objetivo del proyecto

Visto así, el proyecto funciona como una implementación de uso del paper. El interés principal no está en detallar cada celda del cuaderno, sino en entender cómo se organiza la ejecución del método, cómo se conectan los componentes ya entrenados y cómo se da al usuario la posibilidad de definir el estilo mediante ejemplos visuales. El resultado es una versión operativa de la propuesta del artículo, centrada en la transferencia de estilo a partir de referencias seleccionadas por el propio usuario.

Como referencia complementaria dentro del contexto de la asignatura, resulta también interesante mencionar otros trabajos recientes en los que se emplean modelos neuronales para resolver o aproximar problemas matemáticos. En particular, el benchmark de redes neuronales informadas por la física y redes de Kolmogorov-Arnold propuesto en [1] ilustra otra línea actual de aplicación de herramientas de aprendizaje profundo a problemas de carácter matemático y computacional.

Capítulo 2

Modelo base y arquitectura general

2.1. Núcleo del modelo

La arquitectura sobre la que se apoya el proyecto es la que describe el paper [2]. En su núcleo aparece un *Sparse Autoencoder* entrenado sobre representaciones visuales de tipo CLIP obtenidas a partir de un codificador congelado [3]. El papel de este SAE es descomponer una representación densa en un conjunto de conceptos más dispersos e interpretables. Gracias a ello, el estilo de un artista puede describirse como el subconjunto de conceptos que se activan de manera recurrente en varias de sus obras. Esa es la base teórica que permite construir un perfil de estilo interpretable en lugar de depender de una representación opaca.

2.2. Esquema general

Sobre esa base, el paper plantea una pipeline de transferencia de estilo en tres momentos [2]. Primero se entrena el SAE sobre datos artísticos y generales. Después se codifica un perfil de estilo usando un conjunto de obras de referencia. Por último, ese perfil se utiliza para guiar la generación de una nueva imagen. Nuestro notebook no reproduce la primera de estas fases, que es precisamente la parte más costosa del trabajo original, sino que reutiliza el resultado ya obtenido por los autores para centrarse en las dos fases de uso: construcción del perfil y aplicación del estilo.

2.3. Componentes reutilizados

La implementación trabaja, por tanto, con un modelo base ya entrenado y con una arquitectura ya definida. Para ello clona el repositorio oficial de *LouvreSAE*, instala las librerías necesarias y carga desde Hugging Face los pesos y configuraciones ya preparados. También emplea la infraestructura generativa asociada a *Kandinsky 2.2* [4], que en el paper aparece como modelo imagen a imagen de referencia para realizar el *steering*. En consecuencia, no se entrena un nuevo SAE, no se ajusta de nuevo el codificador visual y no se construye una pipeline generativa distinta, sino que se reutiliza una solución ya completa y se pone en funcionamiento en un entorno accesible.

2.3.1. Qué se mantiene del trabajo original

La implementación conserva las piezas esenciales de la propuesta del paper [2]:

- el SAE artístico ya entrenado;
- el espacio de representación asociado al codificador visual;
- el procedimiento de construcción del *style profile*;
- el mecanismo de *steering* dentro de la pipeline generativa.

2.3.2. Qué no se vuelve a entrenar

Desde el punto de vista práctico, esto significa que el notebook no repite la fase de entrenamiento del artículo. El valor del proyecto está en la integración y en el uso del sistema, no en rehacer su aprendizaje desde cero. Esa decisión es coherente con el espíritu del paper, que precisamente presenta su método como una alternativa ligera frente a procedimientos de *fine-tuning* más costosos [2].

Capítulo 3

Funcionamiento de nuestra implementación

3.1. Entradas del sistema

La implementación desarrollada en el notebook se centra en la parte del proceso que depende del usuario. En lugar de imponer un estilo predefinido, el cuaderno permite construir el estilo de destino a partir de ejemplos visuales concretos. Para ello se distinguen dos entradas claramente separadas. La primera es un conjunto de imágenes de referencia que representan el estilo que se quiere transferir. Estas imágenes pueden corresponder a un pintor, a una escuela artística o a cualquier conjunto coherente desde el punto de vista visual. La segunda entrada es la imagen de contenido, es decir, la imagen sobre la que se desea aplicar el estilo aprendido a partir de las referencias.

3.2. Construcción del perfil de estilo

Una vez cargadas las imágenes de referencia, el notebook utiliza las funciones del repositorio oficial para obtener los conceptos activados por cada una de ellas. Después calcula un vector de solapamiento que conserva aquellos conceptos que aparecen de forma consistente entre las distintas obras. Esa operación traduce directamente la idea central del paper: el estilo no se identifica con una sola imagen, sino con la parte compartida que persiste a través de varias imágenes semánticamente distintas [2]. El resultado es un perfil de estilo compacto que resume la información visual recurrente del conjunto de referencias y que puede inspeccionarse en términos de conceptos activados.

3.2.1. Papel del usuario en esta fase

La intervención del usuario se concentra sobre todo en la selección de las referencias. La calidad y variedad de esas imágenes condicionan el perfil obtenido, ya que el sistema busca regularidades entre ellas. Por eso, aunque el proceso de cálculo es automático, la definición efectiva del estilo depende de la muestra visual que se proporciona como entrada.

3.3. Aplicación del estilo a la imagen de contenido

Con el perfil ya construido, el notebook aplica el procedimiento de *steering* sobre la imagen de contenido. En esta etapa no se entrena nada nuevo; simplemente se toma la representación latente de la imagen que se quiere transformar y se modifica mediante el residuo asociado al perfil de estilo. La generación final se realiza usando la pipeline ya implementada en la librería, con parámetros de inferencia como la intensidad del efecto, la

semilla aleatoria y el número de pasos. De este modo, nuestra implementación conserva la estructura del paper pero la convierte en una herramienta manejable: el usuario elige las referencias de estilo, elige la imagen de contenido y el sistema ejecuta automáticamente la transferencia sin necesidad de reentrenar el modelo base.

3.4. Resultados obtenidos

Los resultados obtenidos muestran que la implementación reproduce de forma convincente la idea central del proyecto: extraer un perfil de estilo a partir de varias imágenes de referencia y reutilizarlo para transformar nuevas imágenes de contenido. En términos generales, el sistema conserva la estructura global de la escena original, pero modifica la paleta cromática, el tratamiento de las superficies, la definición de los contornos y, en algunos casos, también la composición visual. Esto hace que el resultado no sea una simple aplicación de filtros, sino una reinterpretación pictórica coherente con las referencias utilizadas para construir el vector de estilo.

3.4.1. Retratos y escenas humanas

En los ejemplos de la Figura 3.1 se observa que, al aplicar un mismo perfil de estilo, en este caso asociado a Vincent van Gogh, sobre una misma fotografía y variar únicamente el valor de α , la identidad general de la escena se mantiene, pero cambia con claridad la intensidad con la que el estilo modifica la imagen. Como α pondera la suma en el espacio latente antes de la reconstrucción final, valores más altos hacen que el resultado se aleje más de la fotografía original y adopte con mayor fuerza la paleta, la textura y la simplificación formal del estilo aprendido. En cambio, con valores más moderados de α , siguen siendo reconocibles rasgos concretos como la postura, la vestimenta o el ordenador, aunque ya reinterpretados dentro del lenguaje visual del artista de referencia. Esta comparación muestra con claridad que el sistema no solo transfiere un estilo, sino que también permite regular cuánto pesa ese estilo sobre el contenido.



(a) Primera transferencia sobre retrato ($\alpha = 1$). (b) Segunda transferencia sobre retrato ($\alpha = 2$). (c) Tercera transferencia sobre retrato ($\alpha = 3$).

Figura 3.1: Tres resultados obtenidos al aplicar el mismo perfil estilístico sobre una misma imagen de contenido con distintos valores de α .

Los retratos también permiten apreciar que la agresividad del estilo influye directamente en el equilibrio entre fidelidad al contenido e intensidad artística. En la Figura 3.2, por ejemplo, se muestran resultados asociados a estilos muy marcados. En primer lugar, el estilo cubista de **Pablo Picasso** transforma de manera muy fuerte la geometría facial y reduce la importancia del parecido fotográfico, priorizando la fragmentación del plano visual y una ruptura de la perspectiva mediante estructuras angulosas. A continuación, el estilo de **Andy Warhol** manifiesta una simplificación cromática de alto contraste y el uso de paletas saturadas, imitando el aspecto visual de la serigrafía industrial, donde predominan los colores planos y la repetición sobre la profundidad tonal. Por otro lado,

la transferencia basada en **Claude Monet** proyecta una atmósfera etérea donde las pinceladas breves y yuxtapuestas diluyen los bordes para capturar la cualidad vibrante de la luz. Finalmente, en el caso de **Vincent van Gogh**, predominan formas más naturales y en movimiento, con trazos curvos que transmiten una sensación de dinamismo y emoción sobre la imagen original. Este conjunto de ejemplos es relevante porque muestra que el sistema no se limita a recolorar la imagen, sino que puede trasladar propiedades estructurales y texturas complejas cuando estas forman parte del perfil aprendido.

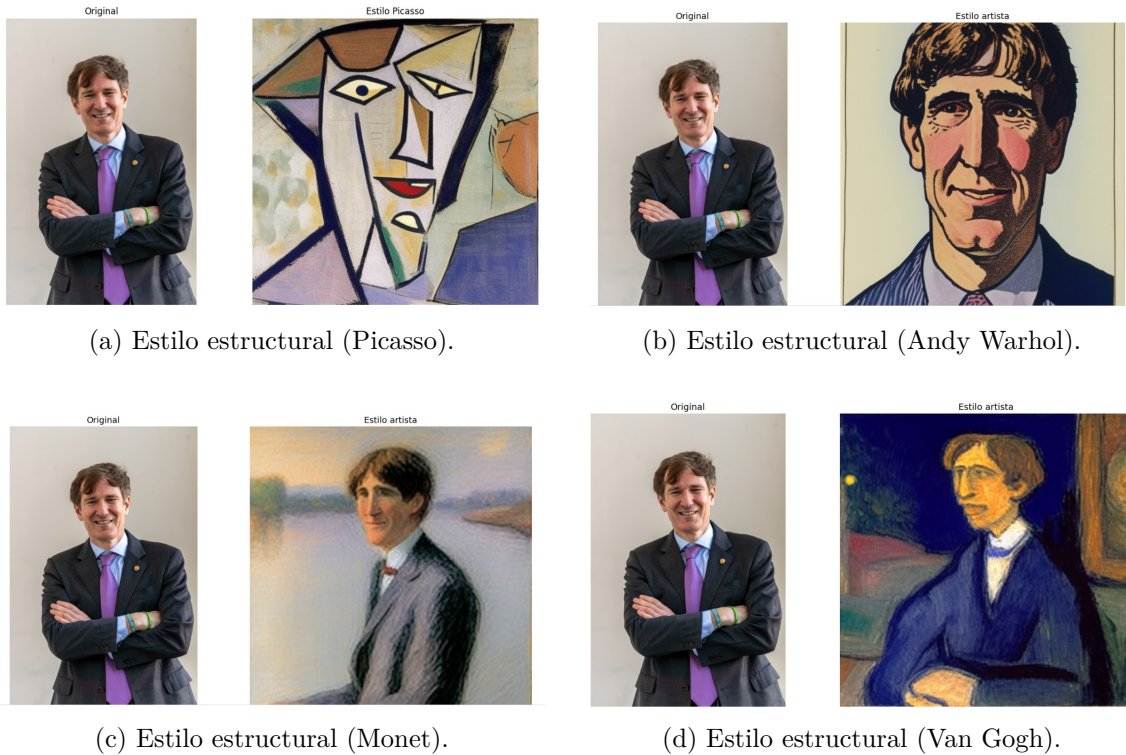


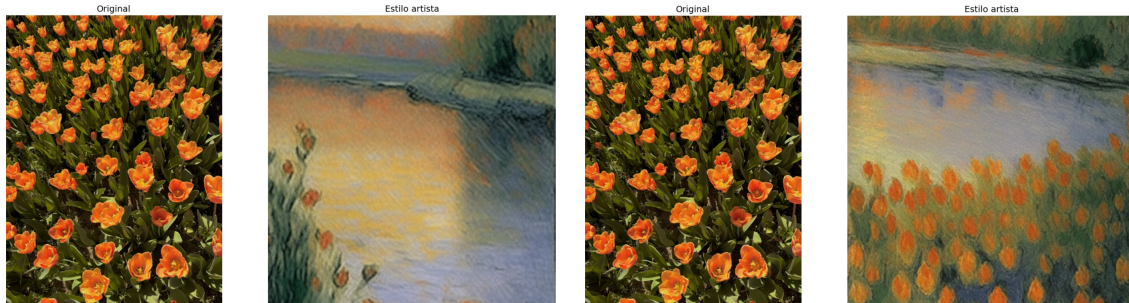
Figura 3.2: Comparativa de transferencia de estilo en diferentes configuraciones de red.

Al contrastar estas cuatro aproximaciones, se evidencia que la transferencia de estilo no actúa de manera uniforme sobre la imagen de contenido. Mientras que los estilos de Warhol y Monet tienden a preservar la legibilidad de las facciones originales, el primero mediante la gestión del color y el segundo a través de la atmósfera lumínica, las propuestas de Picasso y Van Gogh ejercen una “agresión” formal mucho más profunda. En el caso de Picasso, la deformación es puramente geométrica y estructural, llegando casi a la abstracción; en Van Gogh, la alteración es textural, donde la fuerza de la pincelada compite con el detalle del retrato. Esta comparación permite concluir que el éxito de la transferencia depende de lo bien que se combine la imagen original con la capacidad del algoritmo para reproducir el estilo de cada autor.

3.4.2. Escenas naturales y objetos

Los ejemplos de la Figura 3.3, cuyo estilo pertenece a **Claude Monet**, muestran un comportamiento algo distinto cuando la imagen de partida no es un retrato, sino una escena natural con muchas repeticiones locales. En ambos casos el campo de tulipanes deja de percibirse como un conjunto de flores individualizadas y pasa a leerse como una masa cromática continua, con predominio de manchas naranjas, verdes y azuladas. Aun así, el sistema conserva el ritmo visual del motivo original y la separación general entre primer plano y fondo, lo que indica que la transferencia estilística sigue apoyándose en la

estructura global de la imagen de entrada. La diferencia entre ambas variantes no proviene de haber cambiado de estilo, sino de haber modificado el valor de α . Al aumentar este parámetro, el paisaje pierde más detalle fotográfico y se vuelve más pictórico y sintético; al reducirlo, la escena conserva en mayor medida la organización original, aunque ya transformada por la paleta y la textura del perfil aprendido.



(a) Primera transferencia sobre paisaje floral ($\alpha = 2$). (b) Segunda transferencia sobre paisaje floral ($\alpha = 1$).

Figura 3.3: Resultados sobre una escena natural, donde el estilo transforma la textura y la percepción del detalle, pero conserva la organización global de la imagen.

A continuación, si desplazamos la mirada hacia una escena de paisaje como la de la Figura 3.4, vemos cómo el algoritmo interpreta un espacio abierto natural utilizando la impronta característica de **Vincent van Gogh**. Lo que antes era una fotografía nítida y definida se convierte aquí en una pieza vibrante que respira el estilo del pintor holandés, reconocible por sus trazos sinuosos y su empaste texturizado, muy cercano al trazo táctil del pastel o la tiza. Es interesante observar cómo el sistema decide simplificar la complejidad de la arena y las hojas de la palmera, dejando de lado el detalle minucioso para centrarse en la energía del color y el movimiento de la mancha rítmica. La figura humana ya no se presenta como un elemento aislado, sino que parece fundirse suavemente con el entorno, donde el azul profundo del mar y la calidez de la orilla crean una atmósfera mucho más emocional y evocadora que la imagen de partida.



Figura 3.4: Reinterpretación artística de un paisaje costero aplicando el estilo vibrante y textural de Vincent van Gogh.

El caso de la Figura 3.5 confirma, además, que el método no queda restringido a retratos humanos o paisajes. Al aplicarse el estilo de **Andy Warhol** a un objeto cotidiano, el sistema genera una versión visualmente mucho más gráfica, con contornos marcados, superficies planas y una expresividad cercana a la ilustración pop. Esto es interesante

porque pone de manifiesto la versatilidad del enfoque: el vector de estilo puede reutilizarse sobre imágenes de naturaleza distinta siempre que el modelo encuentre una estructura suficiente sobre la que proyectar el nuevo lenguaje visual.



Figura 3.5: Transferencia sobre un objeto cotidiano, con un resultado dominado por el contorno, la simplificación gráfica y el color plano.

3.5. Conclusión

En conjunto, estos resultados permiten extraer una conclusión clara: la implementación es capaz de reutilizar de forma efectiva el perfil de estilo calculado a partir de imágenes de referencia y aplicarlo a contenidos muy distintos entre sí. Además, el parámetro α ofrece un mecanismo explícito para regular el equilibrio entre conservación del contenido y transformación artística. Cuando α es más contenido, la escena original sigue siendo más reconocible y la transferencia afecta sobre todo a la textura y a la paleta. Cuando α aumenta, la imagen resultante se aleja más del parecido fotográfico y se acerca con mayor intensidad al lenguaje visual del artista de referencia. Precisamente en esa capacidad de controlar la influencia del estilo es donde se aprecia con más claridad la utilidad del método propuesto en el paper y de la implementación realizada en el notebook [2].

Bibliografía

- [1] Salvador K. Dzimah, Sonia Rubio Herranz, Fernando Carlos Lopez Hernandez, and Antonio López Montes. A unified benchmark of physics-informed neural networks and kolmogorov-arnold networks for ordinary and partial differential equations. *arXiv preprint arXiv:2602.15068*, 2026. doi: 10.48550/arXiv.2602.15068.
- [2] Raina Panda, Daniel Fein, Arpita Singhal, Mark Fiore, Maneesh Agrawala, and Matyas Bohacek. Louvresae: Sparse autoencoders for interpretable and controllable style transfer. *arXiv preprint arXiv:2512.18930*, 2025. doi: 10.48550/arXiv.2512.18930.
- [3] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139, pages 8748–8763. PMLR, 2021. doi: 10.48550/arXiv.2103.00020.
- [4] Anton Razzhigaev, Arseniy Shakhmatov, Anastasia Maltseva, Vladimir Arkhipkin, Igor Pavlov, Ilya Ryabov, Andrey Kuts, Alexander Panchenko, Andrey Kuznetsov, and Denis Dimitrov. Kandinsky: An improved text-to-image synthesis with image prior and latent diffusion. *arXiv preprint arXiv:2310.03502*, 2023. doi: 10.48550/arXiv.2310.03502.