Laboratoire PPSM – UMR CNRS 8531
Photochimie et Photophysique Supramoléculaires et Macromoléculaires
&
Centre Borelli – UMR CNRS 9010

Séminaire : jeudi 8 décembre 2022 – 15h00
Amphitéâtre 1Z77

Prof. Dr. Robert G. Erdmann

University of Amsterdam and Rijksmuseum, The Netherlands

Invité par : Loïc Bertrand et Victor Gonzalez

A joint language-vision-XRF embedding for investigating the microstructure of paintings

“Operation Night Watch” is a multi-year project in the Rijksmuseum centered on Rembrandt’s large 1642 masterpiece *The Night Watch*. The project is divided into a research phase and a conservation phase. In the research phase, whole-painting imaging was collected across a variety of modalities, including visible-light photography (VIS), UV-induced visible fluorescence photography (UVF), reflectance imaging spectroscopy (RIS-VNIR and RIS-SWIR), macro x-ray fluorescence (MA-XRF), and structured-light stereo photography. When stitched, the 8439 individual visible-light captures form a gargantuan 717 gigapixel image covering the entire surface of the painting (17.2 m^2) at a sampling resolution of 5 µm. The combination of high resolution and a wide breadth of imaging modalities provides an unprecedented research tool for study of the painting since it essentially functions as a virtual whole-object microscope. However, the massive size of the data brings a new problem: there are too many pixels for one person to interpret; when one finds an interesting feature in the microstructure there is no way to find matching areas or to relate this to the features from the other modalities to study the connection between structure and composition.

Recent advances in large mixed language-vision models enable simultaneous training of the language and vision halves by utilizing a large corpus of captioned images extracted from the internet as supervision. These systems learn an image-text joint embedding: images and captions are separately embedded in a common high-dimensional latent space such that images and their associated captions are embedded nearby each other. We propose here to utilize a pre-trained CLIP model to embed all of the tiles from our 5 µm-resolution visible image of the *Night Watch*, thereby enabling language-based queries to find microstructural features matching a given description or exemplar-based queries to find other areas that are “semantic matches” to a given microstructure from the Night Watch or other painting in the sense of being well described by the same captions.

The strong connection between the microstructure and the composition of the paint layer, as captured by the MA-XRF scanning, leads us to further propose adding XRF spectra as a “third door” to the portion of the embedding space associated with painting microstructures. Given an XRF spectrum for a location on a painting, this system embeds the spectrum in the image-text joint latent space, thereby enabling the system to propose characteristic microstructures residing nearby in the latent space. The inverse mapping can also be learned: given a very high-resolution microstructural image, the model can estimate the elemental composition that the forward model would embed nearby to the image in the latent space. Applied across an entire painting, the model thus enables estimating MA-XRF maps from high-resolution photography.