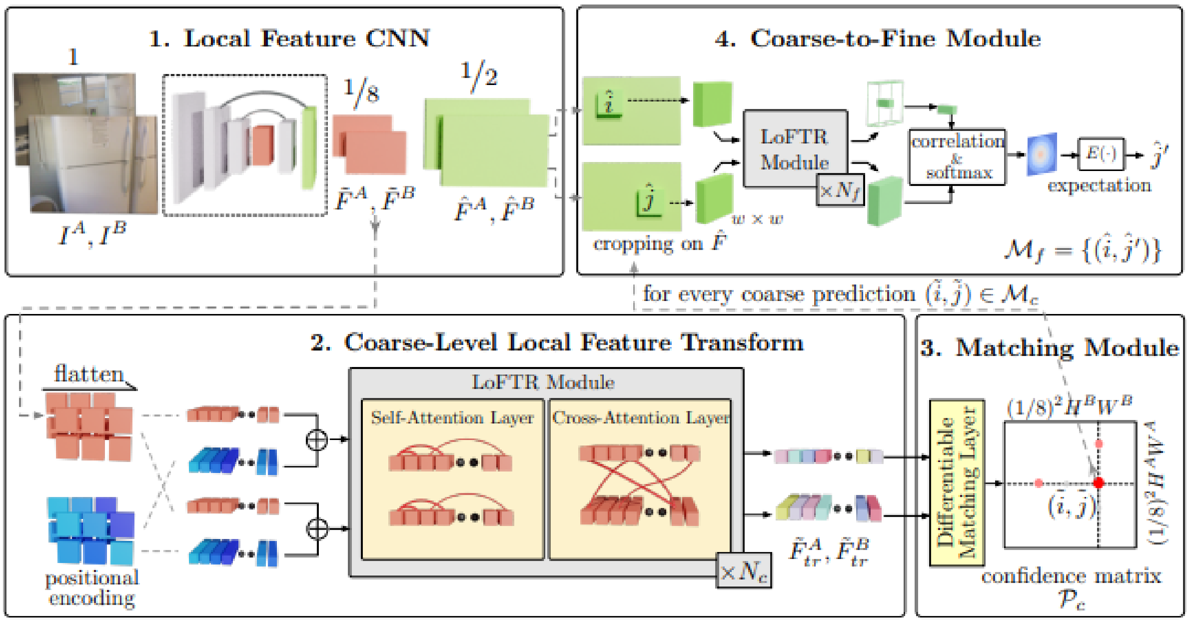
LoFTR for successive brain scan matching

In the literature, a distinction is made between dense and sparse matchers: sparse approaches refer to those with a so-called detection step. This step is intended to find points in the image that are as unique and recognizable as possible (e.g. under a change of perspective). SIFT and SuperPoint are examples of such approaches.
Many of the current sparse image matching methods rely on pretraining based on synthetic data for this detection step: The synthetic data set consists of simple geometric shapes and a detection model is trained to find specific points of these shapes, e.g. corners. In the next step, this detection model is also applied to the original data to find interest points.
Two reasons speak against the use of such an approach: First, synthetic pretraining is a strong prior and it cannot be said with certainty that it will work for any medical images. Secondly, the brain images require as many and uniformly distributed matches as possible. For these requirements, dense matchers, which do not have a detection step, are better suited. One of these approaches is LoFTR:



1: Here, a grid is placed over the two images and each cell in this grid (e.g. 8x8 pixels per cell) is used as an interest point. This is simply done by downscaling the images to ⅛ of their original size.
2: These cells/patches are then flattened so that each patch of the image is represented by a feature vector. Self-attention and cross-attention are used to relate these feature vectors to the other feature vectors, both from the same image and from the other image.
3: In the matching module, the scalar product of each feature vector of image A with each feature vector of image B is calculated. These values result in a score matrix which is converted to a confidence matrix using a dual softmax. Under the mutual neighbor condition, coarse predictions are selected on the confidence matrix (this step can also be understood as patch-classification: Which patch of image A belongs to which patch of image B?)

4: Since the relationship between the points is only known at patch level at this point, this is refined in the last step. To do this, the feature vectors are used at 1/2 the size of the original image (see green features). The matching feature vector of image B to the center feature vector of image A is searched for. Note that this only happens in the patches predicted by the confidence matrix.

Use LoFTR for brain images: The LoFTR paper works with images of size 480x640. This is a huge difference in comparison to the size of the brain images (which are up to 40,000x40,000 pixels). This is intractable using the same approach. But this can possibly be solved using a so-called hierarchical approach: First, two smaller images of the original images are matched. In order to match the original images as well, the entire image no longer needs to be matched at once, as prior information is now available: By matching on the smaller images, it is possible to roughly say which points of the larger images belong together. In order to make the entire procedure as generic as possible, hand-crafted methods (e.g. SIFT) are not used initially.