# Handcrafted Histological Transformer (H2T): A Brief Introduction

**Quoc Dang Vu**[1] QUOC-DANG.VU@WARWICK.AC.UK
**Kashif Rajpoot**[2] K.M.RAJPOOT@BHAM.AC.UK
**Shan E Ahmed Raza**[1] SHAN.RAZA@WARWICK.AC.UK
**Nasir Rajpoot**[1,3,4] N.M.RAJPOOT@WARWICK.AC.UK
[1] *Tissue Image Analytics Centre, Department of Computer Science, University of Warwick, UK*
[2] *School of Computer Science, University of Birmingham, UK*
[3] *The Alan Turing Institute, London, UK*
[4] *Department of Pathology, University Hospitals Coventry & Warwickshire, UK*

**Editors:** Under Review for MIDL 2022

## Abstract

Recently, deep neural networks (DNNs) have been proposed to derive unsupervised WSI representations; these are attractive as they rely less on expert annotation which is cumbersome. However, a major trade-off is that higher predictive power generally comes at the cost of interpretability, posing a challenge to their clinical use where transparency in decision-making is generally expected. To address this challenge, we present a handcrafted framework based on DNN for constructing holistic WSI-level representations.

**Keywords:** Computational Pathology, Unsupervised Learning, Deep Learning, WSI Representation, Transformer

## 1. Introduction

Currently there are two major line of approaches for WSI-level analysis. The first one is to construct features based on classification, detection or segmentation of the tissue components. Despite their effectiveness and capability in providing interpretation, they require a vast amount of annotated samples. The other line has been designed to rely less on such detailed annotations by treating WSI-level tasks as multiple instance learning (MIL) problems. Recent proposals such as CLAM (Lu et al., 2021) employed attention mechanism to enhance the MIL pipeline and achieved notable successes in WSI-level cancer subtyping.

In this paper, we consider the attention mechanism utilized in these approaches as a less powerful version of those from Trasnformer (Vaswani et al., 2017). By using recent knowledge from trying to unveil the power of Transformer mechanism (Ramsauer et al., 2020), we construct a handcrafted version of Transformer, termed as Handcrafted Histological Transformer (H2T) for unsupervised representation of WSIs. We have demonstrated that our handcrafted approximation performed on par with the original Transformer on the cancer subtyping task.

**Disclaimer.** This paper is a short version of our full investigation here (Dang Vu et al., 2022) that is still under peer-review.

## 2. Method

The multi head (self) attention (MHA or MHSA) architecture and its powerful modeling capacity is centered around the following formulation:

$$\widehat{Q} = softmax\left(\frac{1}{\sqrt{d_k}}QK^T\right)V = softmax\left(\frac{1}{\sqrt{d_k}}QW_QW_K^TK^T\right)VW_V \tag{1}$$

Here, $\widehat{Q}$ is the attention output of a single head while $K$, $Q$ and $V$ are commonly referred to as the key, query and value inputs. We denote the associated dimensions of their features as $d_k$, $d_q$ and $d_v$. Additionally, $W_K \in \mathbb{R}^{d_k \times d_e}$, $W_Q \in \mathbb{R}^{d_q \times d_e}$ and $W_V \in \mathbb{R}^{d_v \times d_e}$ are learnable weights for projecting each input feature into a common space with dimensionality $d_e$.

According to (Ramsauer et al., 2020), by using the same input $Y$ for $K$ and $V$ and by renaming the input $Q$ as $R$, Equation (1) can take the form:

$$\widehat{Q} = softmax(\frac{1}{\sqrt{d_k}}QW_QW_K^TK^T)VW_V = softmax(\beta RW_QW_K^TY^T)YW_V \tag{2}$$

where $\beta$ is a scaling factor. By considering $R$ as trainable, we effectively obtain an architecture that learns a set $P$ of prototypical patterns from the training set and how the instances are dynamical weighted against each $p \in P$.

However, $P$ as well as the weight of each instance can be derived effectively without training. We propose using clustering as a mean to derive $P$. With the clustering centroids as histological patterns $p \in P$, we then reformulate Equation (2) into a more generic form

$$\overline{H_i} = \frac{1}{|\Phi_i|}\sum_{\forall \psi_j \in \Phi_i} f(p_i, \psi_j) \odot \psi \tag{3}$$

$$\overline{H} = Concat(\overline{H_0}, ..., \overline{H_N}) \tag{4}$$

with $\psi$ denoting the feature vector of an image patch. Here, $\overline{H_i}$ is representation when projecting the WSI against the $i$-th prototypical histological pattern $p_i$ and $\Phi_i$ is the set of image patches assigned to $p_i$. Specifically, a patch $\psi$ is assigned to a pattern $p_i$ when the distance between their representations is the smallest compared to all other patterns. In Equation (3), $f(p_i, \psi_j)$ is an attribution function that measures the similarity between $p_i$ and $\psi_j$ and $\odot$ denotes the element-wise multiplication of two vectors. For this short paper, we utilized euclidean distance ($\overline{H}$-w) and top-k selection ($\overline{H}$-k128) as $f(p_i, \psi_j)$‘.

## 3. Experiments and Results

We utilized 6 different datasets consisting of a total of 5,306 WSIs from 1,245 patients from The Cancer Genome Atlas (TCGA) and Clinical Proteomic Tumour Analysis Consortium (CPTAC). We focused mostly on the WSIs which were obtained from patients afflicted with either lung adenocarcinoma (LUAD) or lung squamous cell carcinoma (LUSC). Although there are slides that may come from the same patient, for simplicity, in this study we treated each WSI as an independent sample. We alternatively used TCGA and CPTAC as the discovery cohort (training and validation) while using the other as an independent testing cohort. We conducted 5 stratified folds cross-validation on classifying LUAD vs LUSC WSIs and reported results in Table 1.

Table 1: Comparison study on classifying LUAD vs LUSC WSIs. Reported results are mean ± standard deviation of AUROC taken across 5 stratified folds.

| Method | CPTAC-valid | TCGA-test | TCGA-valid | CPTAC-test |
|---|---|---|---|---|
| CLAM(Lu et al., 2021) | 0.9766±0.0054 | 0.8403±0.0033 | 0.9375±0.0065 | 0.9178±0.0031 |
| transformer-1 | 0.9793±0.0073 | 0.8353±0.0083 | 0.9369±0.0110 | 0.9281±0.0063 |
| transformer-2 | 0.9830±0.0079 | 0.8433±0.0052 | 0.9432±0.0119 | 0.9221±0.0026 |
| $\overline{H}$-w | 0.9721±0.0089 | 0.7880±0.0129 | 0.9265±0.0163 | 0.9032±0.0074 |
| $\overline{H}$-k128 | 0.9844±0.0040 | 0.8021±0.0054 | 0.9432±0.0104 | 0.9239±0.0048 |

We used SWAV-ResNet50(Caron et al., 2020) to extract features from image patches where each is of size 512 × 512 at 0.5 micron per pixel. The proposed H2T representations $\overline{H}$ were derived based on 16 prototypical patterns (clustering centroids). These patterns were obtained by using patch-level features extracted from all WSIs within each discovery cohort. $\overline{H}$-w is obtained by weighted summing patch features assigned to a pattern. $\overline{H}$-k128 is obtained by averaging features from the top 128 closest patches assigned to a pattern. We compared these representations against CLAM(Lu et al., 2021) and two version of Transformer: 'transformer-1' based on Equation (2) and 'transformer-2' based on Equation (1). We have demonstrated that our proposal achieved competitive performance on par with the original Transformer counterparts.

## References

Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin. Unsupervised learning of visual features by contrasting cluster assignments. 2020.

Quoc Dang Vu, Kashif Rajpoot, Shan E Ahmed Raza, and Nasir Rajpoot. Handcrafted Histological Transformer (H2T): Unsupervised Representation of Whole Slide Images. *arXiv e-prints*, art. arXiv:2202.07001, February 2022.

Ming Y. Lu, Drew F. K. Williamson, Tiffany Y. Chen, Richard J. Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature Biomedical Engineering*, 5(6):555–570, Jun 2021. ISSN 2157-846X. doi: 10.1038/s41551-020-00682-w. URL https://doi.org/10.1038/s41551-020-00682-w.

Hubert Ramsauer, Bernhard Schafl, Johannes Lehner, Philipp Seidl, Michael Widrich, Lukas Gruber, Markus Holzleitner, Milena Pavlovic, Geir Kjetil Sandve, Victor Greiff, David P. Kreil, Michael Kopp, Gunter Klambauer, Johannes Brandstetter, and Sepp Hochreiter. Hopfield networks is all you need. *CoRR*, abs/2008.02217, 2020. URL https://arxiv.org/abs/2008.02217.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017. URL http://arxiv.org/abs/1706.03762.