EECS 6322 Week 9, Paper 2
Wavelet Flow: Fast Training of High Resolution Normalizing Flows

Shivani Sheth (Student #: 218011783)
shivs29@yorku.ca

I. Summary and contributions

The paper proposes an alternative architecture in normalizing flows, known as wavelet flow. It is a wavelet- based multi-scale normalizing flow architecture that aims to resolve the problem of high training time required by the current architectures in normalizing flow. The proposed architecture is able to develop generative models for high resolution data such as 1024 x 1024 images that was not possible with the previous architectures. It also has an explicit representation of signal scale that includes low resolution signal models and high resolution (super resolution) signal models, and has a competitive performance on low-resolution data with the current state of the art architectures, while having a significantly lower training time.

The proposed architecture is based on wavelets which computes both the probability density and sampling faster, and also decreases the training time for the model, even for high-resolution data. The architecture can be applied to any structured data domain such as audio, video, 3D scans, and images, but the experiments in the paper are restricted to images. Orthogonal wavelets are utilized in the architecture to generate a multi-scale signal representation, which is a convenient method to be used in normalizing flows. The distribution over a wavelet representation of a signal follows a natural coarse-to-fine conditional decomposition, which leads to efficient and parallel training. The authors also introduce a novel sampling architecture to draw samples from an annealed version of the resulting distribution.

Two main concepts required for Wavelet Flow, namely Wavelets and normalizing flows are summarized here. Wavelets are a localized, multi-scale decomposition of a signal. The discrete wavelet transformation that is constructed recursively can be given by $I_{i-1} = h_l(I_i)$ and $D_{i-1} = h_d(I_i)$, where I is an image, D_{i1} are detail coefficients at level $i-1$, h_l applies the wavelet's low-pass filters, and h_d applies the wavelet's high-pass filters. The wavelet transform is also invertible and thus the original image can be recovered. Normalizing Flows are an application of the change of variables formula for probability density estimation. The density function of a random variable X can be given by $p_X(x) = p_Z(f_\theta(x))|detJ_\theta(x)|$, where Z is a random variable with probability density function $p_Z(z)$, with a distribution assumed to be a Normal distribution with mean 0 and unit variance. The function f_θ is invertible and differentiable, parameterized by θ.

The Wavelet Flow can be derived by applying the the change of variables to arrive at a distribution of the wavelet coefficients, which can be given by $p(I) = p(H(I))|det\frac{\partial H}{\partial I}|$. Here, since the wavelets are orthogonal, $|det\frac{\partial H}{\partial I}| = 1$ and hence the product rule of probability can be applied to conditionally factorize the distribution. The Wavelet Flow architecture is trained by maximizing the log likelihood over a set of sample images, where the conditional distribution of detail coefficients for each level can be trained independently. This makes the training process faster and efficient since the computations can be parallelized hence making them easier to fit in limited GPU memory. Sampling from the Wavelet Flow is done recursively where sampling from each distribution is basically sampling from the base distribution of that flow and applying its inverse flow transformation. The architecture uses Markov Chain Monte Carlo (MCMC) separately at each scale to the annealed, unnormalized distribution to sample from an annealed flow with affine couplings and the No-U-Turn Sampler (NUTS) algorithm to generate the samples.

Wavelet Flow uses Adamax optimizer for training and a fully convolutional, and patch-wise training is used for the highest resolution conditional distributions. The distributions were trained using a batch-size of 64 without gradient checkpointing on a single NVIDIA TITAN X (Pascal) GPU. The architecture was preprocessed on datasets such as ImageNet and Large-scale Scene Understanding (LSUN) bedroom, tower, and church outdoor, and trained on high-resolution datasets such as CelebFaces Attributes High-Quality (CelebA-HQ) and Flickr-Faces-HQ (FFHQ). It was compared against baselines such as RealNVP and Glow on the bits per dimensions (BPD) evaluation metric. Wavelet Flow is tested on ImageNet, LSUN, CelebA-HQ, and FFHQ datasets where it performs competitively with the other models on the first two datasets. The results for the last two datasets cannot be obtained by the previous architectures since they are infeasible to compute. The primary achievement of proposed architecture was its training efficiency, which was about 15× and 6× faster than Glow on the first two datasets. The architecture also required a small amount of annealing (T = 0.97) to produce good visual results as compared to Glow which required T = 0.7 for the CelebA-HQ dataset. The architecture also produces high resolution samples trained on the CelebA-HQ and FFHQ datasets at 1024 × 1024 image resolution, which is not feasible

by the previous state of the art architectures.

The conditional structure of a Wavelet Flow also enables it to perform probabilistic super resolution by generating the detail coefficients given a lower resolution input image. Hence, given an input image, the detail coefficients can be generated by sampling from the distribution $p(D_i|I_i)$ and constructing an image to achieve a 2× increase in resolution. This method can be applied iteratively to produce high resolution images. A comparative study between the additive and affine coupling methods also prove that the affine layers generally produce better quantitative results when annealed, as compared to the additive layers. Hence, the architecture uses a single, affine coupling-based model for both quantitative evaluation and sample generation. Overall, although Wavelet Flow has worse performance than the Glow baseline with respect to the spatially distant dependencies in the images, it is able to train the generative model much faster on high resolution data.

II. Strengths

The Wavelet Flow architecture is up to 15× faster to train as compared to its current state of the art architectures in normalizing flows since the architecture is able to efficiently parallelize its training procedures. It also enables the training of generative models for high resolution data which was impractical with previous flow-based approaches, and is the first architecture to work on high resolution data (1024 x 1024 images) in the domain of Normalizing Flows. The multi-scale structure of the Wavelet Flow model can additionally be used to extract consistent distributions of low resolution signals, as well as to perform super resolution.

III. Weaknesses

Qualitatively, the Wavelet Flow architecture has worse performance than the Glow baseline. The spatially distant dependencies in the generated images are not well captured by the architecture and the global coherence of fine details, e.g., eye colour, gaze direction, and hair texture, are inconsistent over larger distances, especially at high frequencies. Additionally, other choices of orthogonal wavelets were not explored in the paper.

IV. Correctness

The claims and empirical methodology of the paper are correct and are supplemented by strong theoretical grounding, related background, and evaluation results from four different datasets namely ImageNet, LSUN, CelebA-HQ, and FFHQ. The model has made it possible to train on high resolution images which was infeasible before. The shortcoming of the model and future work have also been proposed by the authors with respect to the comparative studies with the current state of the art models.

V. Clarity

The paper is clear and well written. The theoretical grounding and the background information provided before the architectural implementation helped in better understanding of the model. The graphs, illustrations, and the comparisons between the current state of the art models also supplemented the ease of understanding. Overall, the paper was also well formatted.

VI. Relation to prior work

Previous architectures in the domain of Normalizing Flows required months of GPU training time to achieve the state of the art results. A few methods proposed earlier to resolve this issue included methods such as multiscale flow and Haar wavelet transformation, both of which limit the complexity of the transformations in certain dimensions but do not explore the natural scale structure of the signal. Wavelet Flow, on the other hand, exposes the scale structure of the signal through a wavelet transformation of the data. The structure of the wavelet can also be found in autoregressive models, but they differ from Wavelet Flows since the conditioning in these autoregressive models is defined by an arbitrary pixel ordering, whereas the conditioning in Wavelet Flow is global but resolution limited.

In the domain of multi-scale image representations, several previous work include models that generate the residual images of a Laplacian pyramid representation that generate an overcomplete representation making it unsuitable for use in a normalizing flow. Some other models also follow a stage-wise multi-scale image generation approach but forego the residual modelling and directly generate higher resolution images given lower resolutions inputs and hence they differ from Wavelet Flow.

VII. Reproducibility

There are enough details provided in the paper including the model implementation (with the code provided on Github), along with the implementation details of the architecture in terms of preprocessing, training, and testing which would help reproduce the major details of the work proposed by the authors.

VIII. Additional comments

Different implementations of Wavelet Flow on architectures such as MaCow, Flow++, and SoS Flow along with different orthogonal wavelets could be extended on the proposed work. The metrics included in the paper could also be briefly explained, along with the highlighted values of the best performance scores.