Fun for the Whole Family: Fast and Furious Transforms

Alex Xie* Alan Hsu[0] Abhishek Vijayakumar
Carnegie Mellon University
`{vroom.vroom, skrt.skrt, nierrrrrrrrrrrr }@andy.cmu.edu`

Abstract

We propose Fast and Furious Transforms *(FFTs), a family of image transforms powered by* ***family****. Our FFT is a two-stage process which transforms an image in the frequency domain into a combination of the most touching family moments in the Fast and Furious movies. Through careful theoretical and empirical analysis, we discover that our transforms are by far one of the transforms of all time. We thus conclude that we don't have transforms – we got family.*

1. Introduction: Family Values

The most important thing in life will always be the people in this room. Right here, right now (emerge). Salute, mi familia.

–Vincent Propane

Modern machine learning research has made great strides in the past decade, but at the cost of the family values we once held dear. To paraphrase one of the great philsophers of our day [7]:

> It seems today that all you see is GANs violence
> in ICML [1] and OwO degeneracy in ICSE [10].
> But where are those good-old fashioned values on
> which we used to rely?
> I was always lucky there was a family transform.
>
> -Sun Tzu
>
> -Peter Griffin

Appalled by this negligence, we seek a return to these good-old fashioned values on which we used to rely. To this end, we draw inspiration from the *Fast and Furious* franchise, well-known around the world for its dedication to family. In this paper, we propose the Fast and Furious transform, the first state-of-the-art image transform that takes into account the contribution of family in image representation learning.

*Has never watched a Fast and Furious movie

[0]Deliberately kept page numbers because he wanted the "funny SIGBOVIK stamp."

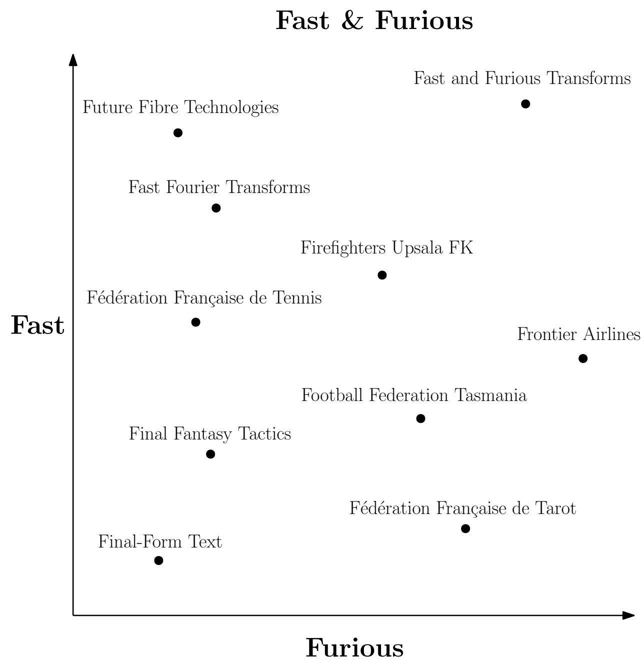

Figure 1: By far one of the figures of all time (in $\log_{10}$ scale).

2. Related Work (hehe do you get it? because family? i hope you got it.)

mu-mu-mu-mu-muridesuu zettai

–Bocchi "The Rock" Johnson [11]

Representation learning: There has been significant interest in recent years in unsupervised representation learning of images [3, 4, 5, 6, 9]. However, these methods fail to account for the inherent familial nature of images. Hence, we group them under the umbrella of **m**ultimodal **i**mage **l**earning **w**ithout **f**amily (MILwF).

FFT: There has also been significant interest in recent years in the acronym "FFT." Fast Fourier Transform (FFT) notwithstanding, we must also compete with *Final Fantasy Tactics* (FFT), Fédération

Figure 2: Sample results for frequency domain decomposition of CIFAR (top row), ImageNet (middle row) and custom (bottom row) images using **family**

Française de Tarot (FFT), Fédération Française de Tennis (FFT), Firefighters Upsala FK (FFT), Football Federation Tasmania (FFT), Federosijuni futboli Toçikiston (FFT), Four Four Two (FFT), Final-Form Text (FFT), Future Fibre Technologies (FFT), Frontier Airlines (FFT). However, we will show that our FFTs are faster and furiouser than all the other acronyms combined (see Figure 1).

3. Methods: Making a family.

Yes, that shit. And I was good in algebra, and like math and shit. And everything else I failed.

–Jesse

3.1. Putting the FFT into FFTs

The first stage of our FFT pipeline is a FFT. Our decision to apply FFT in FFT is motivated by the belief that family transcends mere visual differences. Only by operating in the frequency domain can we understand the true meaning of family.

To begin, we will first define the notion of a set $\boldsymbol{F}$ that we will call a **family**. Specifically, $\boldsymbol{F}$ is the set of movie frames from the Fast and Furious franchise where the cast says the word "family". For a given $\boldsymbol{F}$, we then define $\tilde{\boldsymbol{F}} = \{FFT_{2D}(f) : f \in \boldsymbol{F}\}$ as the set of 2D Fast Fourier transforms (FFTs) of images in $\boldsymbol{F}$, which will serve as the basis for our image decomposition.

Hence, we can define our FFT as a function $f_{\tilde{\boldsymbol{F}}} : \mathcal{I} \to \mathbb{R}^n$ mapping images to n-dimensional embeddings, where n is the cardinality of $\tilde{\boldsymbol{F}}$. In other words, our FFT decomposes each image as a linear combination of FFTs of **family**.

For a given input image $\digamma \in \mathcal{I}$, we first take its FFT to obtain $\tilde{\digamma} = FFT_{2d}(\digamma)$, which will be the input to the second stage of our FFT.

3.2. Putting the Family into FFTs

Now that we have put FFT into FFTs, we now explain the process of putting **family** in FFTs. Given a video sequence $\boldsymbol{V}$ containing all instances of family conversations from the Fast and Furious franchise, we randomly sample ϕ images to form the family $\boldsymbol{F}$:

$$\boldsymbol{F} = \text{sample\_phi}(\boldsymbol{V}). \tag{1}$$

We now construct our corresponding $\tilde{\boldsymbol{F}}$, and crop the center of each image (as a low-pass filter), flatten them to form vectors of length Φ, and concatenate all images column wise to form the $\Phi \times \phi$ **family matrix** $\mathcal{F}$:

$$\mathcal{F} = \text{stack}_1\left(\{\text{flatten}(\text{crop}(\tilde{f})) : \tilde{f} \in \tilde{\boldsymbol{F}}\}\right) \tag{2}$$

Now, given an FFT input image $\tilde{\digamma}$, we resize it such that when flattened it has dimension Φ, which we will call $\tilde{\digamma}_\Phi$. Finally, we can solve for the feature vector $\hat{f}$ that decomposes $\tilde{\digamma}_\Phi$ into our **family**:

$$\mathcal{F}\hat{f} = \tilde{\digamma}_\Phi \tag{3}$$

$$\hat{f} = (\mathcal{F}^T\mathcal{F})^{-1}\mathcal{F}^T\tilde{\digamma}_\Phi \tag{4}$$

4. Experimental Evaluation: Putting our family to the test.

Hey, we do what we do best. We improvise, all right?

–Paul Walker

In figure 2, we show sample image decompositions using our FFT. The results line up with human intuition.

Additionally, for the CIFAR-10 and ImageNet datasets, we train linear classifiers over the representations yielded by our method and the baseline methods. However, we do not report standard error rate as it does not capture the importance of family bonds. We will instead use alternative bond strength metrics.

4.1. Maybe we can CIFAR, but can family make us CIFARther?

We first evaluate our methods on the CIFAR-10 dataset, a standard image classification dataset. Due to the low resolution of images, our frequency domain transformation is invaluable, proving the importance of visual family invariance. We discover that other methods do not in fact CIFAR into the complex tapestry of family, whereas ours allows us to CIFARthest.[1]

	Ionic ↑	Covalent ↑	James[2] ↑
CLIP	256	432	0.025
SimCLR	126	576	0.024
MoCo	173	132	0.024
FFT (legit)	86	666	0
FFT (ours)	2379	49322	0.025

Table 1: Results on CIFAR-10 dataset. Columns 1 and 2 are in units of kilo-JamesBond/mol and column 3 is in units of kilo-JamesBonds.

4.2. I don't got ImageNet, I got ImageFamily.

We further tested the strength of family bonds in our FFT using the ImageNet dataset, but because the dataset has more classes, we decided to use more generalizable bond metrics to quantitatively assess the performance. The Barry-Bonds metric is used as a control group for the experiment.

	Treasury	Municipal	Barry[3]
CLIP	3.41	2.33	8.84×10^{-12}
SimCLR	2.87	2.83	8.84×10^{-12}
MoCo	3.15	2.19	8.84×10^{-12}
FFT (legit)	2.99	2.50	0
FFT (ours)	420.15	105.69	8.84×10^{-12}

Table 2: Results on ImageNet dataset. Columns 1 and 2 are percentages and column 3 is in units of light-years.

[1] How many more times can we use this bad pun? Keep reading and you'll CI how FAR it goes!

[2] Number of James Bond movies at time method was proposed.

[3] Distance Barry Bonds ran for all his home runs at time method was proposed.

As we can see, the treasury and municipal bonds of our FFT are much stronger than those of the rest of the models. We can attribute that to our clever usage of **family** embeddings. Thus, we can conclude that FFT (ours) is stonks.

5. Conclusion: Maybe the real transform was the family we made along the way.

Bing chilling.

–John Cena [2]

We conclusively show that our FFTs outperform competing methods in all relevant metrics (not including accuracy). But at the end of the day, life is not about optimizing performance and winning petty competitions. It's about being fast. It's about being furious.[4] It's about family.

References

[1] Samuel Albanie, Sébastien Ehrhardt, and João F. Henriques. Stopping gan violence: Generative unadversarial networks. In *Proceedings of the 11th ACH SIGBOVIK Special Interest Group on Harry Quechua Bovik.*, 2017. 1

[2] John Cena. Eating ice-cream | bing chilling. https://www.youtube.com/watch?v=AWOyEIuVzzQ, 2021. [Online; accessed 1-April-2023]. 3

[3] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. *arXiv preprint arXiv:2002.05709*, 2020. 1

[4] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey Hinton. Big self-supervised models are strong semi-supervised learners. *arXiv preprint arXiv:2006.10029*, 2020. 1

[5] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. *arXiv preprint arXiv:2011.10566*, 2020. 1

[6] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 1

[7] Seth MacFarlane. Family guy theme song [orginal - hd]. https://www.youtube.com/watch?v=eeNHV2C5m3w&ab_channel=DungeonDogs, 2012. [Online; accessed 1-April-2023]. 1

[8] Tech N9ne, Joey Cool, King Iso, and Dwayne "The Rock" Johnson. Face off. https://www.youtube.com/watch?v=E9T78bT26sk, 2021. [Online; accessed 1-April-2023]. 3

[9] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry,

[4] It's also about drive. It's about power. We stay hungry we devour. [8]

Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision, 2021. 1

[10] Fwans Skawman and OwOlav. Optimal degeneracy through owo based variable names. In *Proceedings of the 16th ACH SIGBOVIK Special Interest Group on Harry Querulous Bovik.*, 2022. 1

[11] Erika Yoshida and Keiichiro Saito. Bocchi the rock!, 2022. 1