Experiments.....

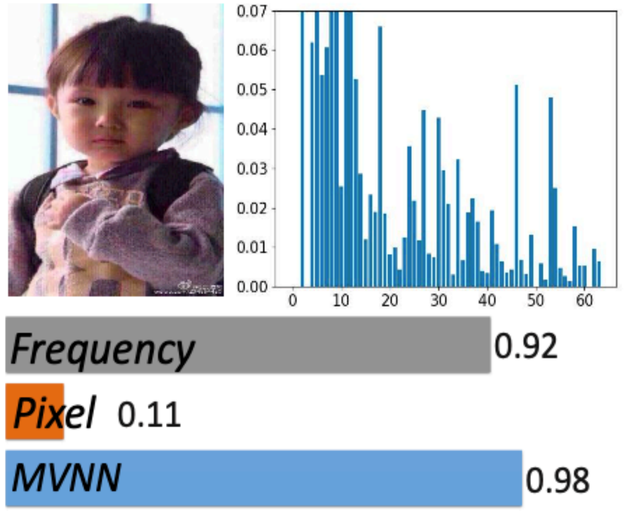
Application on Multi-modal Fake News Detection

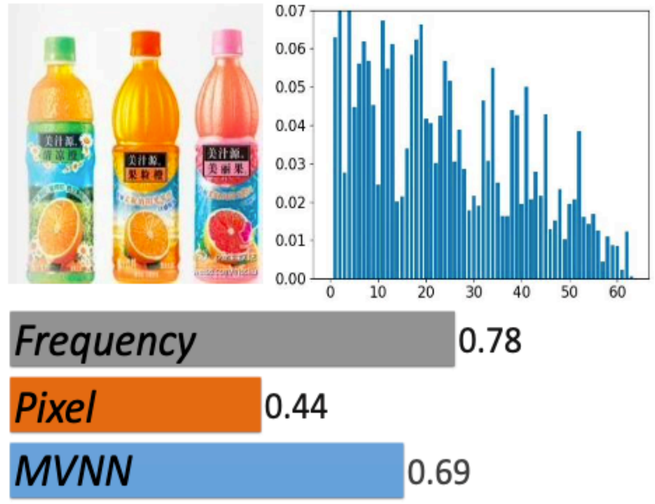
- MVNN consistently outperforms other baselines
  - MVNN by over 5.2% in accuracy, MVNN can easily replace existing methods to obtain the representations of visual contents
- FF+LR in attRNN is obviously worse than EANN and MVAE
- attRNN can hardly utilize the semantic alignment between the text and the forensics features to fuse the textual and visual information

	Method	Accuracy	Precision	Recall	F1
attRNN	FF+LR	0.735	0.801	0.665	0.727
	Pre-trained VGG	0.821	0.813	0.862	0.837
	Fine-tuned VGG	0.849	0.888	0.818	0.852
	ConvAE	0.816	0.848	0.796	0.821
	MVNN	**0.901**	**0.911**	**0.901**	**0.906**
EANN	FF+LR	0.780	0.840	0.724	0.778
	Pre-trained VGG	0.821	0.861	0.791	0.824
	Fine-tuned VGG	0.841	0.883	0.807	0.843
	ConvAE	0.823	0.863	0.794	0.827
	MVNN	**0.897**	**0.930**	**0.872**	**0.900**
MVAE	FF+LR	0.777	0.776	0.815	0.795
	Pre-trained VGG	0.813	0.893	0.737	0.804
	Fine-tuned VGG	0.832	0.875	0.798	0.835
	ConvAE	0.827	0.831	0.847	0.839
	MVNN	**0.891**	**0.896**	**0.898**	**0.897**

Case Studies

captured by MVNN
but missed by pixel domain





- **Frequency:** their frequency histograms look quite suspicious, showing that they are very likely to be **outdated images**.
- **Pixel:** the two images do not show evidence of fake news
- These two examples are correctly classified by MVNN while the results would totally change without the frequency information