Understanding Cross-Cultural Visual Food Tastes with Online Recipe Platforms

Qing Zhang
Chair for Information Science,
University of Regensburg,
Germany
Qing.Zhang@ur.de

Christoph Trattner
Information Science &
Media Studies Department,
University of Bergen,
Norway
Christoph.Trattner@uib.no

Bernd Ludwig
Chair for Information Science,
University of Regensburg,
Germany
bernd.ludwig@ur.de

David Elsweiler
Chair for Information Science,
University of Regensburg,
Germany
david.elsweiler@ur.de

Abstract

Traces of human behaviour with online recipe portals offer an opportunity to employ a data-driven approach to the study of food culture. Here, we focus on understanding visual aspects of food preference by analysing datasets from China, Germany, and US. Predictive modelling with low-level image features and Deep Neural Network image embeddings show differences in recipe images across datasets and between recipes with high and low appreciation within datasets. Our findings demonstrate the utility of the approach for studying visual aspects food culture.

Introduction and Motivation

In nutritional anthropology it is well established that food is much more than mere sustenance and serves diverse needs from health and well-being to control, social contact and ritual (Anderson 2014). Decisions regarding what we eat are complex, not to mention context- and culturally- dependent (Bellisle 2005). As such, significant effort has been undertaken in diverse fields to understand food choices.

Humans have extremely varied diets as they adapt to their environment. This is illustrated by comparing the Inuit diet, consisting nearly exclusively of meat and fats, to that of farmers in South-East Asia, which contains almost no animal protein at all (Fischler 1988). Yet, environment alone cannot explain diet, which is an "evolutionary product of environmental conditions and of the basic forces, especially social institutions and social relations, that determine their use" (Harris and Ross 2009). Explaining food choices requires a blend of biological and cultural factors.

One cultural aspect, which can help explain what we eat as well as how much, is varying aesthetic ideals (Palmer and Schloss 2010; Taylor, Clifford, and Franklin 2013). For food, aesthetics relate primarily to taste (Sherman and Billing 1999), smell (Rolls 2005; Ehrlichman and Bastone 1992) and visual appearance (Linné et al. 2002; Spence et al. 2016). While some preferences are widespread, such as the taste for spicy, herbal and floral volatile oils (Sherman and Billing 1999), geographical differences do exist. For example, in Sherman and Billing's study of typical recipes from different countries, the meat dishes analysed originating from African and Asian countries all featured at least one spice and often a combination of many, whereas in Scandinavian countries, one-third of the recipes used no spices at all. While geographically related cultural differences in eating habits have been extensively studied (see e.g. (Anderson 2014, Ch.12)), studies of cultural differences in aesthetic aspects of food choices are limited.

Anthropologists traditionally learn about food habits using expensive qualitative approaches e.g. (Farquhar 2002). In the Digital Humanities, digital or digitized resources are exploited using techniques from computer science, which allow the qualitative approaches traditionally employed in the humanities (close-reading) to be complemented with quantitative tools, enabling patterns to be unearthed in much larger samples or collections of interest (distant-reading) (Moretti 2005). Such digital methods have been applied to online sources, such as traces from food portals e.g. (Wagner, Singer, and Strohmaier 2014), which have provided insights into the food choices people make including how choices are influenced by temporal factors (Kusmierczyk, Trattner, and Nørvåg 2015; Wagner, Singer, and Strohmaier 2014), gender (Rokicki et al. 2016), geographical location (Wagner, Singer, and Strohmaier 2014; Laufer et al. 2015) and social relations (Rokicki, Herder, and Trattner 2017), as well as how preferences vary with incidence of nutrition related disease (Trattner, Parra, and Elsweiler 2017). Moreover, these studies help with the development of food recommendation systems, which have been touted as a powerful weapon against health problems, such as obesity etc. (Trattner and Elsweiler 2017).

As collections contain both images and appreciation data for recipes, it should be possible to mine insight on visual aspects of food preferences. In this preliminary work we investigate the feasibility of the approach by comparing the visual properties of recipes sourced from three large recipe portals from China, Germany and the United States. Concretely, we answer the following research questions:

- *RQ1.* Is it possible to predict the origin of a recipe based only on the visual properties of the associated image?
- *RQ2.* Is it possible to predict the appreciation of a recipe based only on visual properties of the associated image?
- *RQ3.* Are the same predictive features useful across collections?

We end our paper with suggestions for how we can work together with anthropologists and other humanists to improve understanding of how visual aesthetics impact food choice.

Collections

As a data basis for our analyses we source three collections based on content and user interaction data from popular recipe portals in China, Germany and the US. The Chinese dataset was established by crawling the website Xiachufang.com and contains images for 25,508 recipes. The American data were sourced from Allrecipes.com and consists of images for 35,501 recipes. Kochbar.de was the source of the German data, where we obtained images for 72,899 recipes. In all cases we had 1 image for each recipe taking the first, default image associated with each.

Features

In our analyses we make use of two types of features representing diverse aspects of an image's visual properties. The first, which we refer to as Explicit Visual Features (EVF), relates to 10 low level image properties originally proposed by (San Pedro and Siersdorfer 2009). These include the image Brightness, Sharpness, Contrast, Colorfulness, Entropy, RGBContrast, Variation in Sharpness, Saturation, Variation in Saturation and Naturalness. These features have been shown to have utility when predicting the popularity of online recipes (Trattner, Moesslang, and Elsweiler 2018), as well as online food choices (Elsweiler, Trattner, and Harvey 2017). To calculate these features we followed the detailed instructions provided by (Trattner, Moesslang, and Elsweiler 2018) using the OpenIMAJ Framework[1].

A second type of visual feature used is Deep Neural Network image embeddings (DNN). For each recipe image in the datasets we obtain features from a VGG-16 DNN, a convolutional deep neural network developed to classify images (Simonyan and Zisserman 2014). The output is a vector of 4,096 dimensions. In particular, we use a VGG-16 model pre-trained with the ImageNet dataset (Krizhevsky, Sutskever, and Hinton 2012). A vector of 4,096 dimensions was generated with the Keras[2] framework. DNN features calculated in this way have proved to be powerful in a food image retrieval setting (Salvador et al. 2017).

Predicting Recipe Source Collection

Comparing the distributions of the EVF features statistically reveals significant differences across the collections. The only comparison not found to be highly significantly different ($p < .001$) was sharpness when comparing the Kochbar and Allrecipe images. The strongest effect sizes are found when comparing the features for Xiachufang with those of Kochbar (r_{mean} = .16, r_{max} = .23). The effect sizes when comparing Kochbar and Allrecipes are smallest (r_{mean} = .09), but even in this case two features hinted at effects (Saturation r = .17 and Naturalness r = .13)

To establish whether these differences and the estimated DNN features are sufficient to distinguish between the collections, we formulated a prediction task whereby classifiers were trained to predict the source dataset for each image in a random sample of 25,000 images from each collection. We tested the mean accuracy of Naive Bayes (NB), Logistic Regression (LOG) and Random Forest (RF) classifiers using the EVF, DNN and a combined feature set. The results using a 5-fold cross validation protocol are shown in Table 1.

Table 1: Results for predicting which collection an image belongs to based on different feature sets. Best performing scores for each classifier are bolded.

Features	Mean Accuracy		
	NB	LOG	RF
EVF(Brightness)	.41	.41	.42
EVF(Sharpness)	.41	.41	.43
EVF(Contrast)	.37	.37	.43
EVF(Colorfulness)	.38	.38	.40
EVF(Entropy)	.38	.38	.40
EVF(RGBContrast)	.38	.38	.40
EVF(Sharpness Variation)	.41	.41	.42
EVF(Saturation)	.39	.39	.40
EVF(Saturation Variation)	.38	.38	.41
EVF(Naturalness)	.38	.38	.40
EVF(All features)	.47	.54	.55
DNN	.66	**.86**	.78
EVF+DNN	**.67**	**.86**	**.79**

It is clear from the bottom 3 rows that images from different collections are sufficiently visually distinct such that they can be classified with reasonable accuracy, regardless of the classifier employed. The EVF feature set offers the lowest accuracy while the DNN embeddings offer more predictive power. Combining the EVF and DNN sets provides little improvement. The EVF features taken individually perform only slightly better than random (33.3%), but when combined using the RF a far higher accuracy can be achieved.

Predicting Appreciated Recipes

As a next step, we formulated a two-class prediction experiment to determine whether sufficient visual differences exist to distinguish between recipes deemed as "appreciated" or "less appreciated" by the respective communities. For each collection we use the most appropriate and available metric as discussed in the literature (Trattner and Elsweiler 2017). For Xiachufang this was aggregated user rating, whereas for Allrecipes and Kochbar this was log-transformed bookmark count for each recipe. We drew a sample of 5,000 images for each collection (2,500 from the top-10% and 2,500 from the bottom-10% based on the appreciation metric). This is nearly all of the Xiachufang recipes in these percentiles and an undersample of the other two collections, which are larger, with the aim being to draw a fair comparison. The prediction task was performed on each dataset individually. The results, again using a 5 fold cross-validation protocol, are shown in Table 2.

The results suggest that using the visual features to determine appreciation is a harder task than predicting the source collection. Despite having only two classes, the prediction accuracies achieved are lower overall. There is, however, some evidence of signal. The highest performance was attained on the Allrecipes collection and the poorest on Xiachufang. As with the first task, the EVF features offered less predictive power than DNN in all three collections. Interestingly, the best performing EVF feature differs for each col-

[1]http://openimaj.org/

[2]https://keras.io/

Table 2: Results for prediction experiment where the aim was to classify recipes as appreciated (recipe appeared in top 10% of scores) or unappreciated (bottom 10%). Best performing scores for each classifier are bolded.

Xiachufang	Mean Accuracy			**Allrecipes**	Mean Accuracy			**Kochbar**	Mean Accuracy		
Features	NB	LOG	RF	Features	NB	LOG	RF	Features	NB	LOG	RF
EVF(Brightness)	.56	.54	.54	EVF(Brightness)	.59	.58	.58	EVF(Brightness)	.53	.53	.53
EVF(Sharpness)	.59	.58	.59	EVF(Sharpness)	.53	.51	.53	EVF(Sharpness)	.53	.55	.54
EVF(Contrast)	.55	.55	.54	EVF(Contrast)	.52	.52	.52	EVF(Contrast)	.51	.51	.51
EVF(Colorfulness)	.53	.53	.51	EVF(Colorfulness)	.53	.50	.53	EVF(Colorfulness)	.54	.55	.53
EVF(Entropy)	.51	.49	.53	EVF(Entropy)	.54	.54	.54	EVF(Entropy)	.54	.54	.53
EVF(RGBContrast)	.55	.55	.54	EVF(RGBContrast)	.51	.52	.51	EVF(RGBContrast)	.51	.49	.49
EVF(Sharpness Variation)	.59	.60	.59	EVF(Sharpness Variation)	.52	.52	.52	EVF(Sharpness Variation)	.52	.53	.54
EVF(Saturation)	.56	.57	.56	EVF(Saturation)	.55	.52	.53	EVF(Saturation)	.56	.56	.56
EVF(Saturation Variation)	.51	.50	.50	EVF(Saturation Variation)	.52	.51	.51	EVF(Saturation Variation)	.53	.54	.52
EVF(Naturalness)	.54	.54	.54	EVF(Naturalness)	.51	.51	.52	EVF(Naturalness)	.53	.53	.53
EVF(All features)	**.60**	**.63**	.63	EVF(All features)	.58	.60	.61	EVF(All features)	.57	.58	.58
DNN	**.60**	.60	**.65**	DNN	**.64**	**.65**	**.71**	DNN	**.64**	**.62**	**.68**
EVF+DNN	**.60**	.60	.62	EVF+DNN	**.64**	**.65**	.70	EVF+DNN	**.64**	**.62**	**.68**

(a) *Xiachufang:* High (↑) prediction scores

(b) *Xiachufang:* Low (↓) prediction scores

(c) *Allrecipes:* High (↑) prediction scores

(d) *Allrecipes:* Low (↓) prediction scores

Figure 1: Sample of images with high and low prediction scores in Xiachufang (a & b) based on all EVF features, Allrecipes (c & d) based on DNN features.

lection with Sharpness and Sharpness Variation most useful in Xiachufang, Brightness for Allrecipes and Saturation in Kochbar. The combined EVF models come close to achieving the DNN performance on Xiachufang, but on the other datasets DNN performed better by a considerable margin.

Figures 1 (a & b) illustrate the output of a RF model with EVF features on the Chinese collection, which performs relatively well (see Table 2). From these figures, it is relatively obvious that Xiachufang users perceive high contrast images with black backgrounds and white plates positively[3] whereas yellowy, brown images are perceived negatively. This is a finding known in the anthropology literature (Palmer and Schloss 2010). The DNN models, despite offering higher predictive accuracy, are more difficult to interpret

[3]black background images were sourced from different users

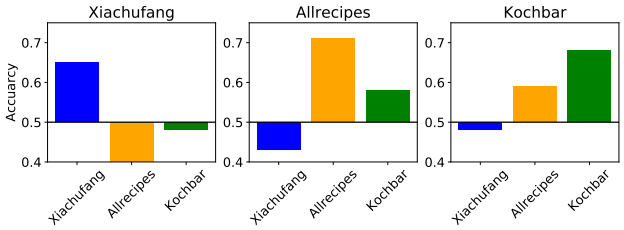

Figure 2: The best performing model for each collection and their performance on the other two collections.

(see Figures 1 (c & d)[4]). This has also been reported in the past (Montavon, Samek, and Müller 2018) and interpreting the patterns will require closer collaboration with anthropologists (see below).

To understand if the same signals are present across collections, we use the best performing model for each collection and test their ability to make predictions for the other sets. Figure 2 shows the results. The Xiachufang model performs well on the Chinese recipes but very poorly (worse than random) on the other collections. The Kochbar and Allrecipes models, in contrast, both perform best on their own images, perform reasonably well on the other collection, but perform poorly on the Xiachufang images. This may be a sign that the visual preferences of German and US users are more similar than for these users and users of Xiachufang.

Limitations, Conclusions and Future Work

When summarising the findings, their meaning and how these can be built on in future work, it is important to acknowledge limitations of the work. One such limitation is that we compare visual properties of images from heterogeneous communities in large countries, such as China, US and Germany and treat them as if they were mono-cultures. This is overly simplistic as shown by past work demonstrating geographical trends in Chinese (Zhu et al. 2013), US (Trattner, Parra, and Elsweiler 2017) and German language (Wagner, Singer, and Strohmaier 2014) food portals. Nevertheless, the results of our simple experiments show that:

[4]A lack of space prevents more examples being shown, but the other DNN models were similarly difficult to interpret.

- Automatically extracting image features can be used to reliably differentiate the photographs posted to online recipes portals in these different countries, meaning that despite the previously evidenced geographical differences within the countries, there is still detectable variation between portals from these countries.
- The same visual features can be used to classify more and less appreciated recipes in all 3 collections, albeit with slightly poorer accuracy than in the first experiment. This aligns with past work (Elsweiler, Trattner, and Harvey 2017), but we show stable patterns across food cultures.
- The Allrecipes model can make useful predictions for Kochbar and vice-versa, but neither model makes good predictions for Xiachufang. Similarly, the Xiachufang model makes poor predictions for both Allrecipes and Kochbar, hinting that German and US visual tastes are more similar to each other than to those of the Chinese.

All three findings suggest value in data-drive approaches as a means to study cultural differences in the visual aspects of online food interactions. What our analyses do not do, however, is explain *how* tastes differ. This was underlined by the difficulty in interpreting the output of the best performing DNN models, but this is a known problem for such models (Montavon, Samek, and Müller 2018). In our future work, we will work closely with anthropologists to address this problem. Our plan is to employ various cluster analyses to determine similar images within different classes (e.g. appreciated or not within datasets) and allowing colleagues to interpret the groupings. The idea would be to receive feedback that would enhance our understanding of the underlying behavioural patterns and, at the same time, improve our methodological approach.

References

Anderson, E. N. 2014. *Everyone eats: Understanding food and culture*.

Bellisle, F. 2005. The determinants of food choice. *EUFIC Review* 17(April):1–8.

Ehrlichman, H., and Bastone, L. 1992. Olfaction and emotion. In *Science of Olfaction*. 410–438.

Elsweiler, D.; Trattner, C.; and Harvey, M. 2017. Exploiting food choice biases for healthier recipe recommendation. In *SIGIR*, 575–584.

Farquhar, J. 2002. *Appetites: Food and Sex in Post-socialist China*.

Fischler, C. 1988. Food, self and identity. *Information (International Social Science Council)* 27(2):275–292.

Harris, M., and Ross, E. B. 2009. *Food and Evolution: Toward a Theory of Human Food Habits*.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. In *NIPS*, 1097–1105.

Kusmierczyk, T.; Trattner, C.; and Nørvåg, K. 2015. Temporality in online food recipe consumption and production. In *WWW*, 55–56.

Laufer, P.; Wagner, C.; Flöck, F.; and Strohmaier, M. 2015. Mining cross-cultural relations from wikipedia: A study of 31 european food cultures. In *WebSci*, 3.

Linné, Y.; Barkeling, B.; Rössner, S.; and Rooth, P. 2002. Vision and eating behavior. *Obesity research* 10(2):92–95.

Montavon, G.; Samek, W.; and Müller, K.-R. 2018. Methods for interpreting and understanding deep neural networks. *Digital Signal Processing* 73:1–15.

Moretti, F. 2005. *Graphs, Maps, Trees: Abstract Models for a Literary History*.

Palmer, S. E., and Schloss, K. B. 2010. An ecological valence theory of human color preference. *PINAS* 200906172.

Rokicki, M.; Herder, E.; Kuśmierczyk, T.; and Trattner, C. 2016. Plate and prejudice: Gender differences in online cooking. In *UMAP*, 207–215.

Rokicki, M.; Herder, E.; and Trattner, C. 2017. How editorial, temporal and social biases affect online food popularity and appreciation. In *ICWSM*, 192–200.

Rolls, E. T. 2005. Taste, olfactory, and food texture processing in the brain, and the control of food intake. *Physiology & Behavior* 85(1):45–56.

Salvador, A.; Hynes, N.; Aytar, Y.; Marin, J.; Ofli, F.; Weber, I.; and Torralba, A. 2017. Learning cross-modal embeddings for cooking recipes and food images. In *CVPR*, 619–508.

San Pedro, J., and Siersdorfer, S. 2009. Ranking and classifying attractiveness of photos in folksonomies. In *WWW*, 771–780.

Sherman, P. W., and Billing, J. 1999. Darwinian gastronomy: Why we use spices: Spices taste good because they are good for us. *BioScience* 49(6):453–463.

Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Spence, C.; Okajima, K.; Cheok, A. D.; Petit, O.; and Michel, C. 2016. Eating with our eyes: From visual hunger to digital satiation. *Brain and cognition* 110:53–63.

Taylor, C.; Clifford, A.; and Franklin, A. 2013. Color preferences are not universal. *Journal of Experimental Psychology: General* 142(4):1015.

Trattner, C., and Elsweiler, D. 2017. Investigating the healthiness of internet-sourced recipes: Implications for meal planning and recommender systems. In *WWW*, 489–498.

Trattner, C.; Moesslang, D.; and Elsweiler, D. 2018. On the predictability of the popularity of online recipes. *EPJ Data Science* 7(1):20.

Trattner, C.; Parra, D.; and Elsweiler, D. 2017. Monitoring obesity prevalence in the united states through bookmarking activities in online food portals. *PloS One* 12(6):e0179144.

Wagner, C.; Singer, P.; and Strohmaier, M. 2014. The nature and evolution of online food preferences. *EPJ Data Science* 3(1):38.

Zhu, Y.-X.; Huang, J.; Zhang, Z.-K.; Zhang, Q.-M.; Zhou, T.; and Ahn, Y.-Y. 2013. Geography and similarity of regional cuisines in china. *PloS One* 8(11):e79161.