Abstract
Prior work has shown that combining deep learning techniques with a massive amount of paired recipe/image data yields impressive results in an image-recipe retrieval task and enables semantic vector arithmetic in the space of food recipes and images. Our demo system showcases how arithmetic operations in a joint recipe/image embedding space create new possibilities for intuitive human engagement with a large and high-dimensional dataset.

Author Keywords
Deep learning; semantic arithmetic; recipes; cooking; artificial intelligence; intelligent interfaces

ACM Classification Keywords
H.5.2 [Information Interfaces and Presentation]: User Interfaces; H.3.4 [Information Storage and Retrieval]: Systems and Software; I.2.6 [Artificial Intelligence]: Learning (Deep Learning); J.m [Computer Applications]: Miscellaneous

Introduction
Access to large datasets via a knowledgeable human (e.g., a chef) allows for a rich vocabulary of partial and relative queries (e.g., “I like lemon cake, but I’m looking for something a bit sweeter.”). Supporting this kind of query using a computer requires inference over several content dimensions (ingredients, form, taste, etc.), but using recent deep...
learning techniques, it is possible to represent food recipes and concepts using vectors, and to describe queries using arithmetic on the representations of concepts (e.g., ‘lemon cake’ + 0.2 * (‘sugar’ - ‘vinegar’); see [1]). Semantic vector spaces derived from complex data using deep neural networks thus present the opportunity to develop a new kind of query interface that is potentially more powerful and intuitive for non-expert users than existing techniques. We have created a publicly available dataset of recipe/image pairs, as well as a derived semantic vector space, and a demo which illustrates some new ways of accessing the data.

System Design
The technical details of our system are outlined in [2]. We gathered more than one million recipes (ingredient lists plus preparation steps) along with images of prepared food (this dataset is available at http://im2recipe.csail.mit.edu/). We trained an image encoder network to produce an encoded vector from the image, a recipe encoder network to produce one from the recipe, and trained both to produce identical vectors for each image-recipe pair. By applying vector arithmetic and finding the recipe or image encodings closest to the result, we can answer semantic queries (this process can also help find, e.g., a healthier analogue). Because human queries are rarely specific recipes, a simple text search of recipe contents to match a query like “Chinese” followed by an averaging step can produce a vector approximation of an arbitrary query (this is one area for future study). Our demo provides a simple web interface for using a few fixed formulae that enable queries which would be difficult to approximate using existing recipe search engines.

Use Cases
Analogy—Using the formula \( R = A - B + C \), analogical reasoning can be simulated by subtracting one component and then adding another. For example, the analogy ‘wrap’ is to ‘chicken quesadilla’ as ‘soup’ is to ‘...?’ can be represented as ‘?’ = ‘chicken quesadilla’ - ‘wrap’ + ‘soup.’ Because cooking instructions are included, this formula supports finding alternate preparations (e.g., ‘french fries’ - ‘fried’ + ‘mashed’).

Interpolation—The space between two recipes can be explored using the formula \( R = xA + (1 - x)B \). Sampling points between two vectors or letting the user actively ‘crossfade’ between them allows for a user-driven discovery process and triangulation of new recipes.

Scale-based Offset—Adding a fixed constant to the interpolation formula gives \( R = C + xA + (1 - x)B \), which allows exploring the scale defined by vectors \( A \) and \( B \) starting from point \( C \). For example, if \( A \) is ‘sugar’ and \( B \) is ‘vinegar,’ the scale should form a sweet/sour axis which can be used to explore variations on any base dish of interest \( C \).

Human-Recipe Interaction
Our system represents an important first step towards semantic-arithmetic-based interfaces for recipe exploration, and we encourage others to work with our data set to further expand this design space. We expect that there are more subtle and powerful ways to leverage this kind of semantic querying for intuitive user interface design, and we are interested in potential collaboration around this topic.

REFERENCES