Final Defense Presentation

MS THESIS
FINAL DEFENSE

FOOD CALORIE MEASUREMENT USING EFFICIENT CLASSIFICATION AND DISTANCE MEASUREMENT TECHNIQUES

BY

NAUMAN ZAFAR

NUST 2014 64096 MSEECS 61314F

PANEL OF ADJUDICATORS

SUPERVISOR

Engr. Dr. Omar Arif

Assistant Professor

COMMITTEE MEMBERS

Dr. Anis ur Rahman

Assistant Professor

Engr. Dr. Faisal Shafait

Associate Professor

Dr. Muneeb Ullah

Assistant Professor

INTRODUCTION

- Nutrients give our bodies instructions about how to function.
- If we don't get the right instructions, our metabolic processes suffer and our health declines.
- “Food acts as medicine--to maintain, prevent, and treat disease”*
- In short, what we eat is central to our health.

*Lines from a Food Journal published by university of Minnesota

- **60 nations***
  Across the world

- **30,000+ consumers**
  Concerned about healthier eating habits

- **41% Generation Z (1990-present)**
  Willing to pay for healthier products

*Report conducted by the Nielsen Global Health and Wellness Survey

SYSTEM DESIGN GOALS

"Food Item Recognition and captioning along with Ingredients and Attributes estimation"

MOTIVATION

- Non or Semi-automatic system
- Cost Effective and Approachable Solution
- Limited Dataset
- Assumptions
- Time Efficiency
- Accurate Gauging of Real Object

KEY CONTRIBUTIONS

- Transfer Learning Based Food Item Recognition along with enhanced Captioning technique
- Extraction of Food Attributes and Ingredients to familiarize user with Food Content
- Introduction of a Food Dataset

PROPOSED SOLUTION

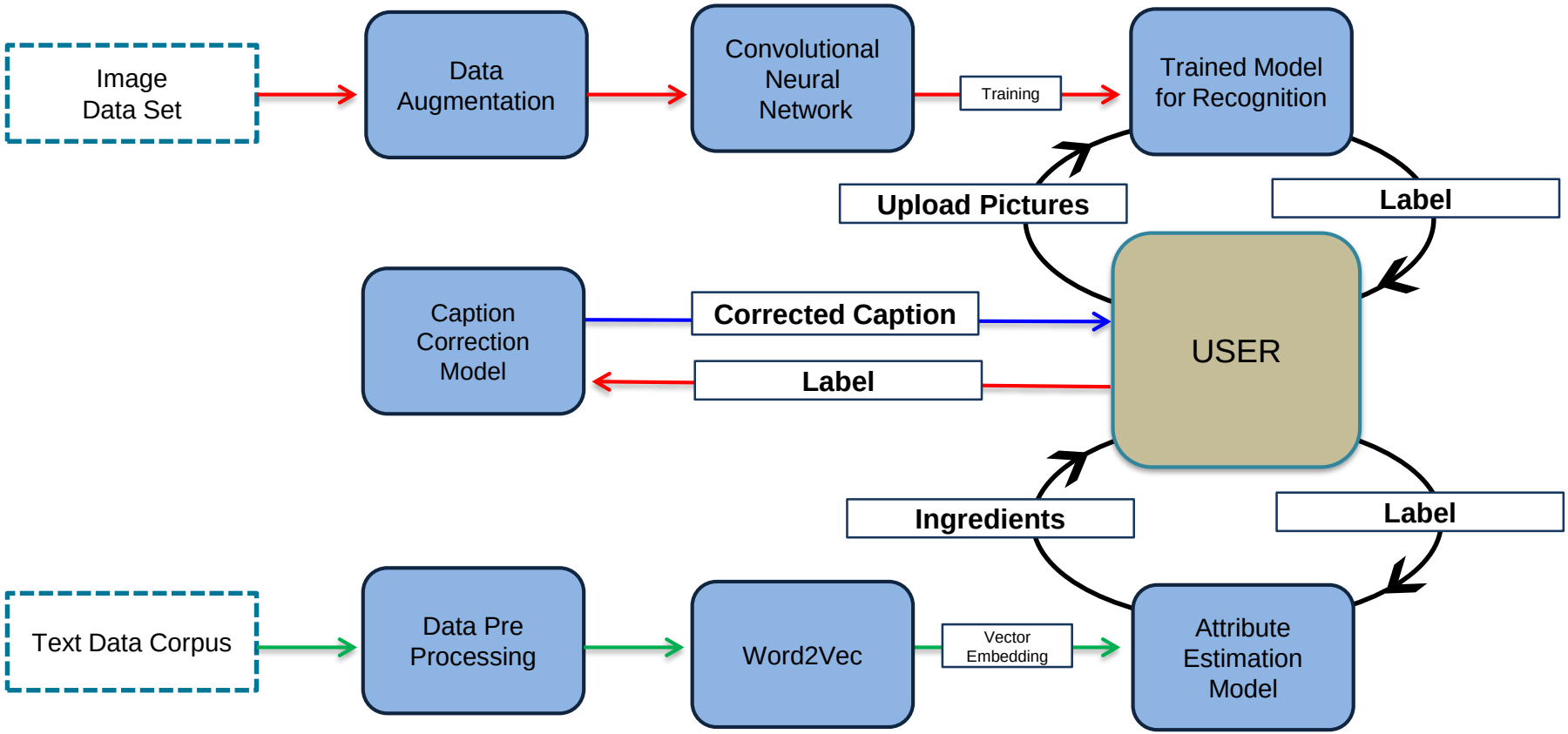
PROPOSED SOLUTION

- Ingestion conscious people can get info about their intake using food image taken from mobile
- BUT HOW?
  - How we know about food image class?
    - Transfer learning based food recognition
  - How to know about food content?
    - Info extraction from text (NLP)
- Text scattered on web related to food items can be beneficial if processed
- BUT HOW?
  - Utilizing Word2Vec

PROPOSED SOLUTION

- More on awareness to diet conscious people
  - Captioning
    - Neural image captioner
    - Not trained on specific food dataset
    - Result: wrong captions for food images
  - We propose Caption Correction

PROPOSED SOLUTION

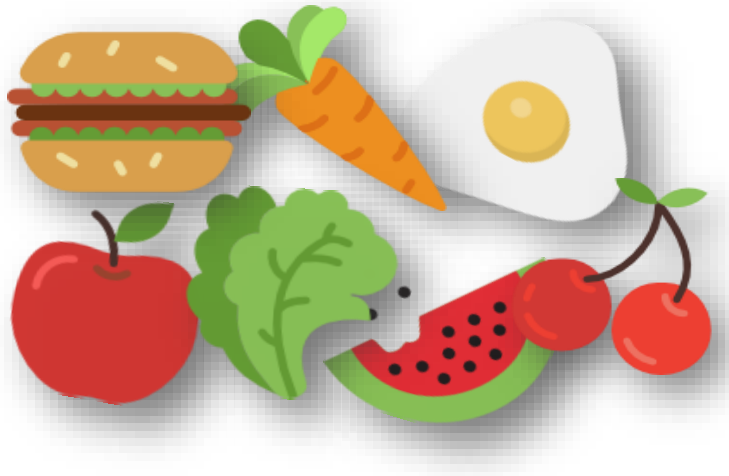


TRANSFER LEARNING BASED FOOD RECOGNITION AND CAPTIONING TECHNIQUE ENHANCEMENT

FOOD IMAGE RECOGNITION MODULE

PRE- TRAINED MODEL ACCURACY IMPROVEMENT

- Food Images Dataset
  - Popular Food Datasets
    - PFID, FOOD-101, TADA
  - In order to get feedback from our regional community,
    - We constitute classes of our regional food.
    - 35% Classes from FOOD-101.
    - Rest of 65% classes is taken from our regional food.
    - 500 to 1000 images in each category.
  - Continental and Asian dishes



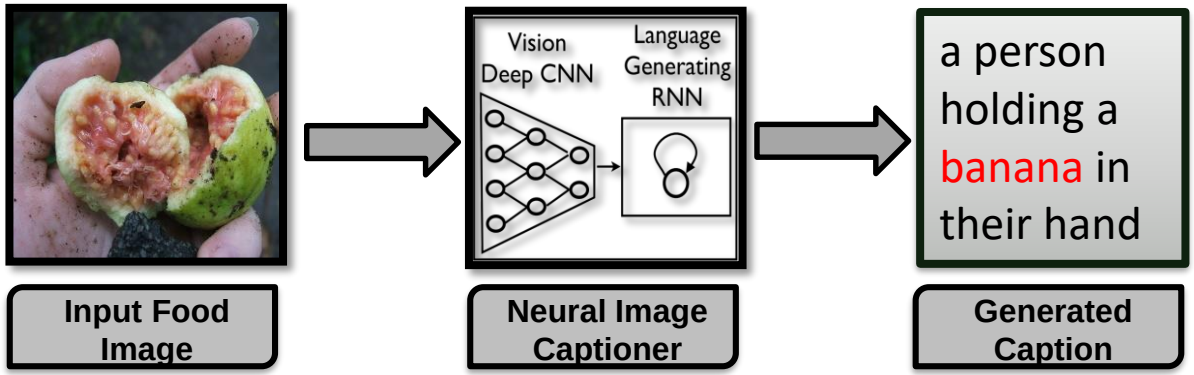
PRE- TRAINED MODEL ACCURACY IMPROVEMENT

- Instead of building model from scratch, we took advantage of a pre-trained model
  - Pre-trained model requires moderate computation with less training time
- Done experiments on different pre-trained models
  - Inception v4 trained with our dataset showed maximum accuracy of 91.73%.
- How finetuned ?
  - Pre-trained weights are loaded
  - FC and Last two conv blocks are trained
- Other improvement approaches
  - Data Augmentation
  - Batch normalization
  - Regularization

FOOD IMAGE CAPTION CORRECTION MODULE

Show and Tell: A Neural Image Caption Generator*

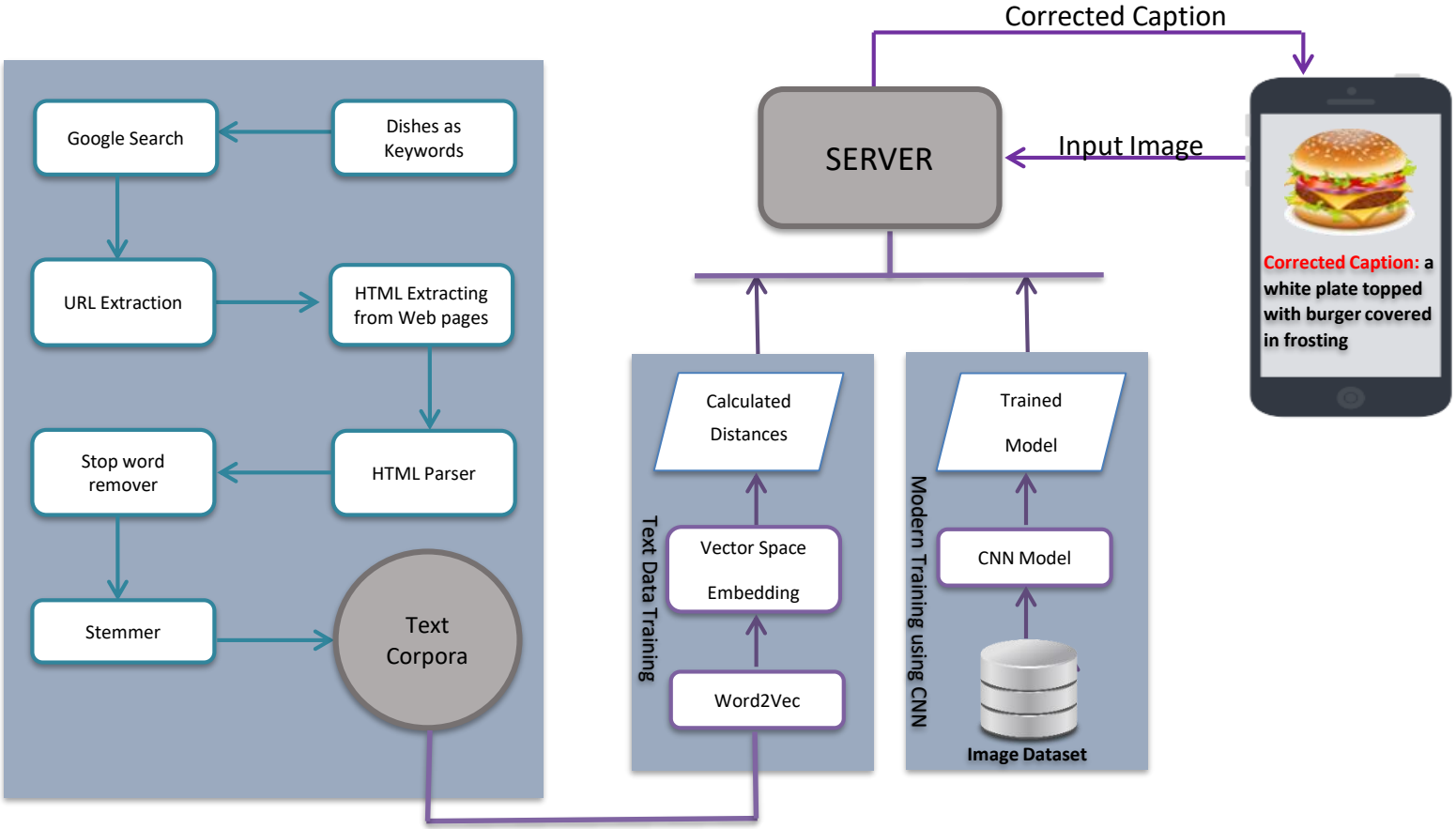
- A generative model which utilizes deep recurrent architecture and used to produce natural sentences describing a picture



- Wrong image captions due to lack food specific classes in dataset
- How to solve?
  - Caption Correction Technique

*Oriol Vinyals et al

WHAT WE PROPOSED : CAPTION CORRECTION TECHNIQUE



PRE-TRAINING PHASE : TEXTUAL DATASET PREPARATION

- Textual Data Congregation
  - For the collection of data, we used two different frameworks. i.e. Common Crawl & Scrapy
  - Captions are generated for all the images containing in our dataset using neural image captioner and all the generated captions are saved in textual form
- Textual Data Cleansing and Pre-Processing
  - HTML to text and removal of Java script tags and programming language code using beautiful soup
  - For Stop words removal, Lemmatization and Stemming, we utilized libraries from NLTK.

TRAINING PHASE : TEXTUAL MODEL TRAINING

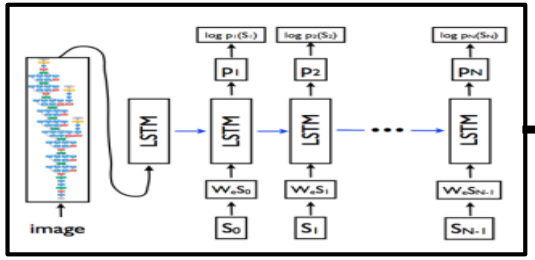
Word2Vec Model Training

- Semantic learning framework and uses a shallow neural network
- Word2Vec takes the whole text corpus as an input, creates a vocabulary of words used in that text
- Overall computes vector space embedding of given text data.
- Used CBOW approach
- After experimenting with different configurations, the most relevant results are achieved using the values described in Table

NAME	VECTOR DIMENSIONALITY	WINDOW	SAMPLE	NEGATIVE	ITERATION
VECTOR MODEL	300	12	1e-4	25	25

EXAMPLE DEPICTING DIFFERENT WORKFLOW PHASES

Food Image



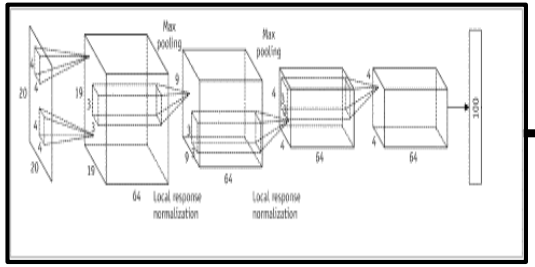
Neural Image Captioner Model

a person holding a **banana** in their hand

Generated Caption

- person
- hold
- banana
- hand

Pruned Caption



CNN Based Recognition Model

Guava

Predicted Label

.527 **banana**

.395 hold

.078 person

Similarity Score Calculation

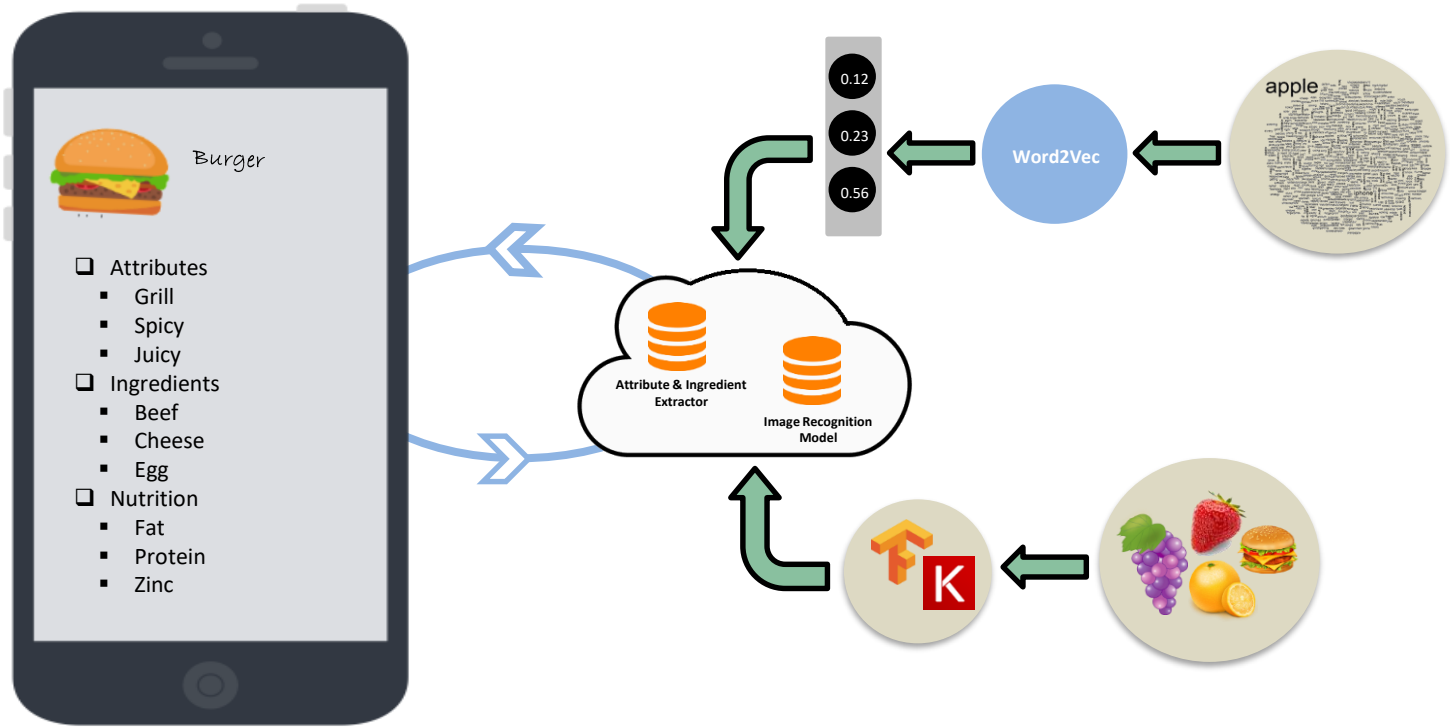
SCORE BASED COMPARISON

a person holding a **guava** in their hand

Corrected Caption

EXTRACTION OF FOOD ATTRIBUTES AND INGREDIENTS

WHAT WE PROPOSED



PRE-TRAINING PHASE : ATTRIBUTE AND INGREDIENTS LIST PREPARATION

- A static list of attributes and ingredients
- Attributes → Food Traits
- Ingredients → Elements present in food item
- Encompassed all possible and relevant attributes in the list

TRAINING PHASE : TEXTUAL MODEL TRAINING

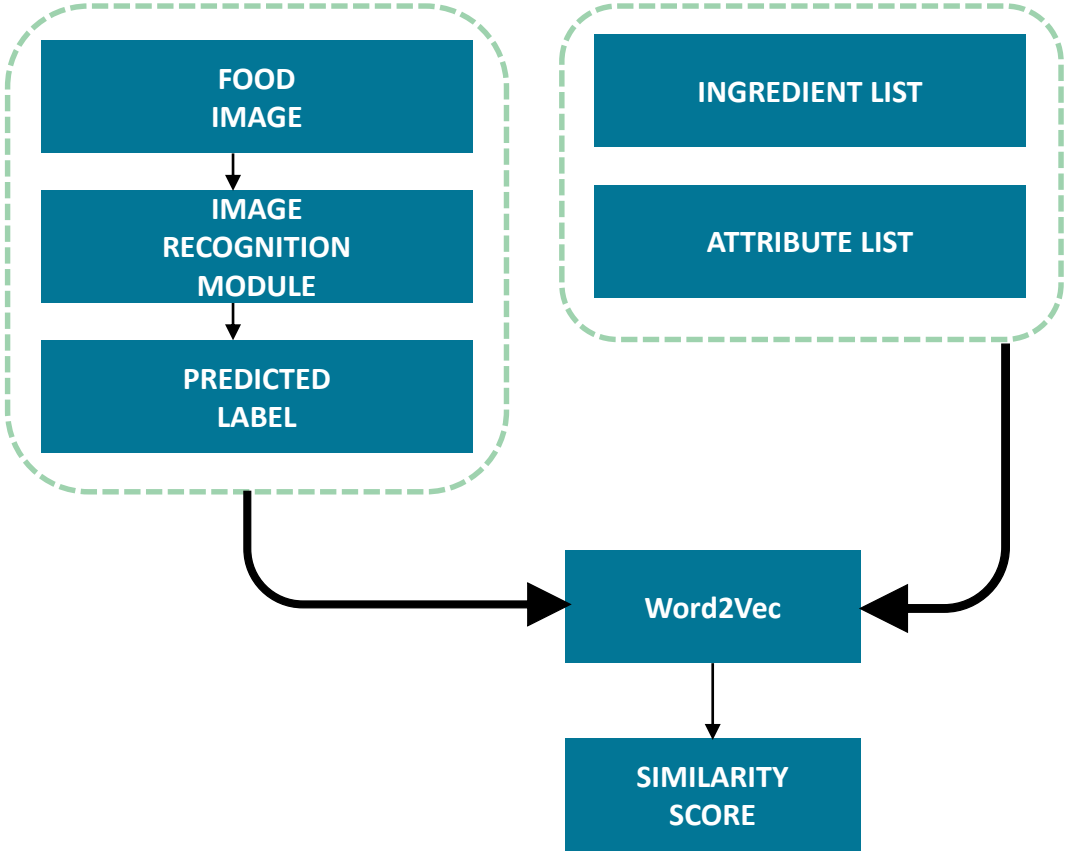
Word2Vec

- Used CBOW approach

NAME	VECTOR DIMENSIONALITY	WINDOW	SAMPLE	NEGATIVE	ITERATION
VECTOR MODEL	200	8	1e-4	25	15

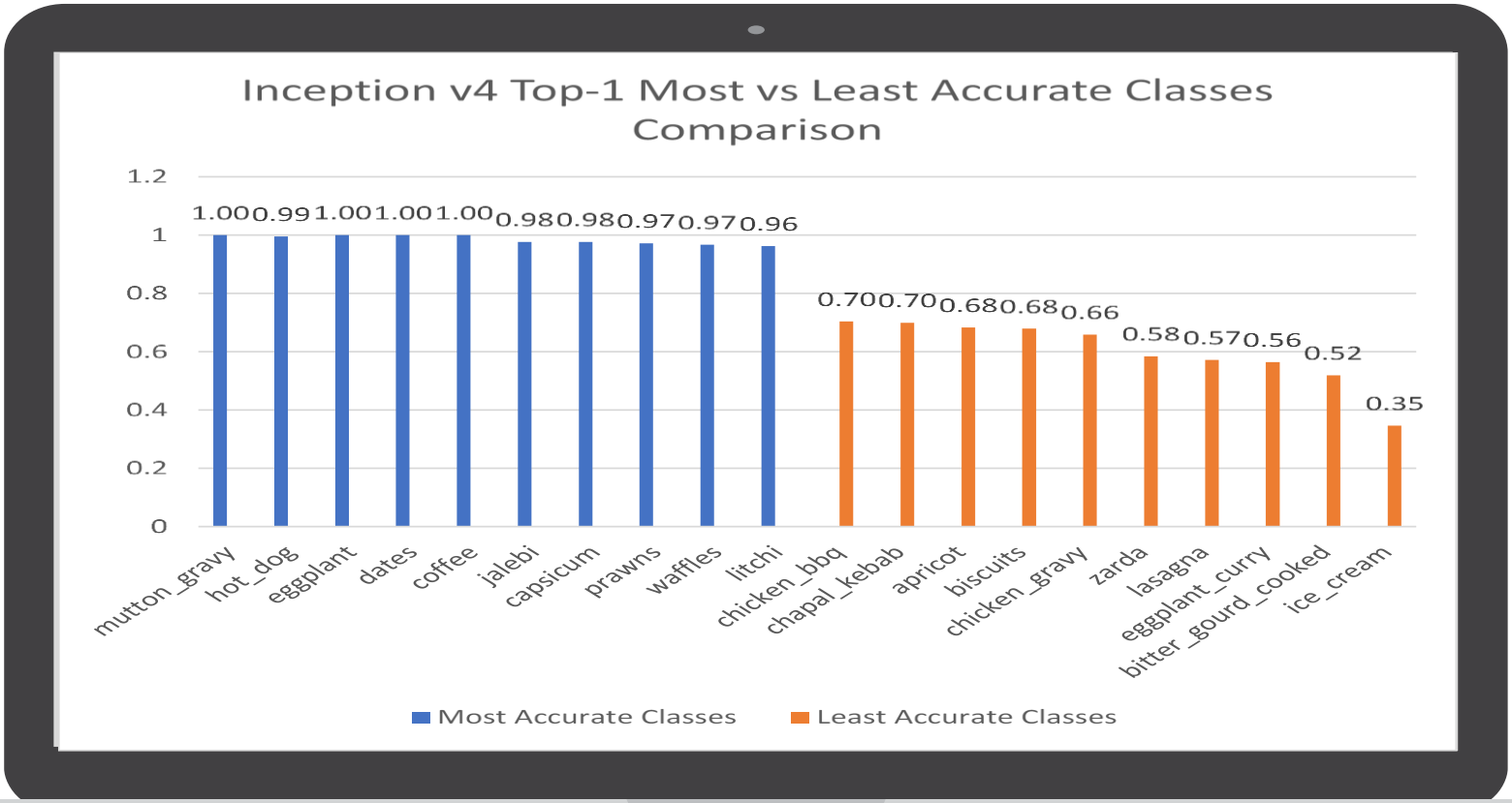
WORK FLOW OF ATTRIBUTE AND INGREDIENTS EXTRACTION

- Ingredients
  - Highest frequency
  - Sample size is large
  - Most accurate
- Nutrition values
  - Sparse In the text
  - Characteristics occur very infrequently
  - Less Accurate



RESULTS AND EVALUATION

INCEPTION V4 TOP-1 MOST VS LEAST ACCURATE CLASSES COMPARISON



More Details 1

CAPTION CORRECTION MODEL RESULTS

Predicted Label: **Tea**
Similarity Score:
{'**food': '0.279'**,'bowl': '0.246','sit': '0.133'}
NIC: a bowl of **food** sitting on a table
Novel: a bowl of **tea** sitting on a table

Predicted Label: **Prawns**
Similarity Score:
{'**sheep': '0. 257**', 'white': '0. 126', 'black': '0. 083'}
NIC: a close up of a white and black **sheep**
Novel: a close up of a white and black **prawns**

Predicted Label: **Prawns**
Similarity Score:
{'**broccoli': '0. 482**', 'food': '0.325', 'plate': '0.196'}
NIC: a close up of a plate of food with **broccoli**
Novel: a close up of a plate of food with **prawns**

Predicted Label: **Guava**
Similarity Score:
{'**banana': '0. 527**','hold': '0.395', 'person': '0.078'}
NIC: a person holding a **banana** in their hand
Novel: a person holding a **guava** in their hand

Predicted Label: **Cheese**
Similarity Score:
{'**food': '0.279'**,'bowl': '0.246','sit': '0.133'}
NIC: a bowl of **food** with a spoon in it
Novel: a bowl of **cheese** with a spoon in it

Predicted Label: **Cucumbers**
Similarity Score:
{'**sandwich': '0.280'**,'top': '0.060','sit': '0.052'}
NIC: a cut in half **sandwich** sitting on top of a table
Novel: a cut in half **cucumbers** sitting on top of a table

Predicted Label: **Tea**
Similarity Score:
{'**wine': '0.355**', 'glass': '0.191', 'sit': '0.133'}
NIC: a glass of **wine** sitting on top of a table
Novel: a glass of **tea** sitting on top of a table

Predicted Label: **Cucumbers**
Similarity Score:
{'**carrot': '0.343**', 'wooden': '0.167, 'pile': '0.086'}
NIC: a pile of **carrots** sitting on top of a wooden table
Novel: a pile of **cucumbers** sitting on top of a wooden table

More Details 1

CAPTION CORRECTION MODEL RESULTS

Actual Label: **Avocado**
Predicted Label: **Bitter_gourd**
Similarity Score:
{**'apple': '0.126'**, 'top': '0.110', 'pile': '0.103'}
NIC: a pile of **apples** sitting on top of a table
Novel: a pile of **bitter_gourd** sitting on top of a table

Actual Label: **Falooda**
Predicted Label: **Spaghetti**
Similarity Score:
{**'glass': '0.329'**, 'vase': '0.307', 'flower': '0.290'}
NIC: a **glass** vase with a flower in it
Novel: a **spaghetti** vase with a flower in it

Actual Label: **Kheer**
Predicted Label: **Haleem**
Similarity Score:
{**'cake': '0.442'**, 'birthday': '0.332', 'candle': '0.226'}
NIC: a birthday **cake** with a candle on it
Novel: a birthday **haleem** with a candle on it

Actual Label: **Zarda**
Predicted Label: **Fried_rice**
Similarity Score:
{**'food': '0.254'**, 'white': '0.092', 'plate': '0.031'}
NIC: a white plate topped with a piece of **food**
Novel: a white plate topped with a piece of **fried_rice**

CONTENT EXTRACTION RESULT ON BURGER IMAGE

Attributes		Ingredients		Nutrition	
Grill	1	Beef	1	Fat	1
Spicy	0.75	Cheese	0.9	Sodium	0.55
Juicy	0.64	Lettuce	0.8	Protein	0.44
Fry	0.54	Tomato	0.67	Cholesterol	0.24
Tasty	0.53	Potato	0.58	Carbs	0.7
Healthy	0.33	Soya	0.53	Zinc	0.3
		Egg	0.52	Manganese	0.15
		Onion	0.50		

CONTENT EXTRACTION RESULT ON BIRYANI IMAGE

Attributes		Ingredients		Nutrition	
Fry	1	Rice	1	Carbs	1
Spicy	0.69	Chicken	0.73	Fat	0.8
Grill	0.55	Saffron	0.63	Starch	0.49
Healthy	0.5	Chilli	0.57	Fiber	0.38
Hot	0.43	Ghee	0.55	Sodium	0.36
Healthy	0.33	Soya	0.4	Manganese	0.35
Juicy	0.39	Tomato	0.3	Zinc	0.23
Sour	0.21				

Conclusion

CONCLUSION

- Curiosity is found among people to measure their heaviness and healthy eating in order to avoid over weightiness
- So we developed a system which can aware user about his/her daily intake
- We have done critical analysis of recent studies on accurate calorie estimation
- We successfully utilized transfer learning for food item recognition model, where it can predict food items with high accuracy
- Accurate and cognitive description of food image to assist targeted audience about what they eat
- Our proposed system can help
  - nutritionist for physical and medical treatment of overweight persons
  - normal people can also get the benefit of this system

FUTURE WORK

FUTURE WORK

- The results can be enhanced by:
  - Using even more large text corpus.
  - Increasing number of attributes.
  - Cross validation.
  - Tuning Word2Vec parameters.

PUBLICATIONS

PUBLICATIONS (submitted)

1. O. Arif, H. Afzal, R. Yunus, H. Noor, S. Tazeen, N. Zafar, M. Faisal, R. Nawaz, "Real Time Estimation of Nutritional Value of Food through Attribute Estimation Using Deep Learning and Vector Embedding", *Proceedings of the International Journal of Wireless Communications and Mobile Computing*.
2. M. Muhammad Ali Baig, M. Ihtisham Shah, M. Abdullah Wajahat, N. Zafar, O. Arif, "Image Caption Generator with Novel Object Injection", *Proceedings of IEEE Transactions on Instrumentation & Measurement*.
3. N. Zafar, "Food Item Recognition and Calorie Measurement Techniques: A Review", *Proceedings of the Sukkur IBA Journal of Computing and Mathematical Sciences*.

THANK YOU

ANY QUESTIONS

END