Final Project Report

Le Thi Hong Ha (210205), Nguyen Hoang Ngoc Ha (210206),

Tu Khanh Dang (200022), and Le Duc Dai Loc (210086).

Fulbright University Vietnam

CS101: Computer Science I

Instructor: Dr. Phan Thanh Trung

31st May, 2022

Executive summary

After working on the two videos on Youtube about Vietnamese street food made by chillystudio and FairDinkum Adventure, we extracted the annotated file and see that sometimes, the percentage of humans and food can be smaller than our expectation. We also explored that the feature analysis of the scripts provided us with some word ideas that will attract more audience members with longer viewing times, especially function and verbs.

And the visual feature analysis educated us on the visual elements which contribute to the video's rising popularity in the realm of visuals, especially the humans and food. In this data set, artificial intelligence is able to quickly determine the color of each pixel, the scenes, and humans. The data set used by AI does not contain enough variety to enable it to recognize all of the things in Vietnam, particularly Vietnamese food and locations. By comparing the data retrieved by AI and humans, we are aware of the limits of AI, which will allow us to develop it in the future.

The study of computer science should absolutely include some time spent on learning how to manage projects, including keeping track of the various tasks, deadlines, and assignments we have. We have completed almost all of our projects with the assistance of many project management platforms such as Google Calendar or Doodle, which have been of great assistance to us in meeting all of our deadlines and completing our work on time.

Project background

Needs statement

Street food is a prominent feature of Vietnamese cuisine. It's the most accessible type of food in the country, and it's also the cheapest. This industry in Vietnam has grown significantly in recent years, with an estimated 80% of all restaurants being street vendors. And it is now a significant part of Vietnam’s tourism industry as well, which brings in over 8 million visitors every year. It is not only important because it's inexpensive and accessible to many people - but it also provides a variety of dishes and flavors that are hard to find elsewhere. Moreover, it has become so popular that they have been exported abroad like pho (Vietnamese beef noodle soup). The popularity of these dishes has led to many restaurants serving them internationally as well. While working on this project, we better understood the popularity of Vietnamese street food. Moreover, we can know better how to attract audiences on social media by watching street food and how to use computer science to know the way to spread Vietnamese street food to more people.

Because students in these introductory computer science classes have already been introduced to the idea of video analysis in the very first session of the course, which draws from a range of academic disciplines. This project is centered on the process of extracting information from movies and analyzing what that information tells about the movie. Now we are going to have a look at a couple of videos that have just been uploaded to the channels of chillystudio and FairDinkum Adventure on YouTube. Both channels are owned by FairDinkum Adventure and are about distinct kinds of trips that focus on Vietnamese food and culture.

Goal and objectives

In consideration of skills across a variety of technical domains. There are segmented videos that were analyzed by utilizing the data that was obtained from the string, and it was found that the video exists. Moreover, we have Investigated the outcomes of the AI in conjunction with the annotations that have been added to the various situations. In addition, we are going to carry out an analysis that is descriptive and educate ourselves on how to put our knowledge of the various types of data to use in a scenario that is modeled after the actual world. Additionally, we will find out how the data from YouTube is saved and processed in terms of the visual qualities of the content and think about how the information you obtained in CS1 may be applied to the task of deciphering clips about a variety of various kinds of street food.

As part of our efforts to promote cultural comprehension, we are going to inform viewers about the many benefits that come with consuming street food in Vietnam. In addition, one of our goals is to bring awareness to the necessity of maintaining a healthy diet and to show that food that is bought on the street may be an option that is nutritious. In addition, we would want to encourage viewers to go to Vietnam in order to try this cuisine for themselves by making recommendations that they should do so.

Design constraints and feasibility

The difficulty we faced came not only from maintaining the flow of the movie but also from selecting the appropriate frames for the final product. One of the things that prevented us from doing more was the fact that the movie contains many frames that are less than one second in length. As a corrective measure, we slowed the pace down to a half a frame every second, which gave the impression of greater clarity.

Moreover, one of the most significant difficulties we face is trying to find a balance between the demands of our academic obligations and the demands of the responsibilities we have in other areas of our lives. This is one of the most significant challenges we face. The

solution that we have arrived at is to establish timeframes and adhere to the group meeting directions in the strictest manner that is possible.

In addition to this, the instructions, when seen for the very first time, are extremely complicated and difficult to understand. It is strongly suggested that we ask the professor for some guidance in order to acquire a deeper comprehension of everything that is taking place at this moment.

Literature review

We conduct 5 research in our literature review to support our project report. We understand a conceptual framework to highlight the development areas in the social sustainability of street food in Hanoi (Stutter, 2017). The challenges and social functions that social vendors have to face are also focused on the findings of the research. From the factors of this research, we know the impacts on the development of street food in Hanoi and how they sustain many models of restaurants. On the other aspect, we find that the entrepreneurship factors in the microbusiness of the street food sector in Vietnam (Hiemstra, van der Kooy, and Frese, 2006). We know more about how the street food factor in Vietnam can succeed or not and how the characteristics affect their decision in expanding their business. The key points we gain in reading research are the unclean conditions, and the food safety knowledge of consumers and food vendors in Ho Chi Minh City based on their age, educational level, gender, etc. (Samapundo, Cam Thanh, Xhaferi and Devlieghere, 2016). We understand the favorite aspects of the consumers from the research and the food handling practices of street food vendors. Therefore, we know the impact of how the marketing on street food when focusing on the empirical evidence to help us know the impact of sensory marketing and the revisit decisions to build the marketing strategies of HCM tourism and the degree of influence on street food visitors (Hoang and Tučková, 2021). In addition, the culture, social values, and food are the factors that attract the tourist and highlight the role of the emotion in their answers, and the research we read focused on the emotional responses of international tourists when experiencing street food in Vietnam (Linh, 2021).

From most of the research above, we better knew how the value of Vietnamese street food attracts other people. But this project will aim to analyze how the communication, especially from social media, can use their voice and words to spread the value of street food. Moreover, as a multidisciplinary research field, we will know how to apply computer science in these fields.

Implementation notes

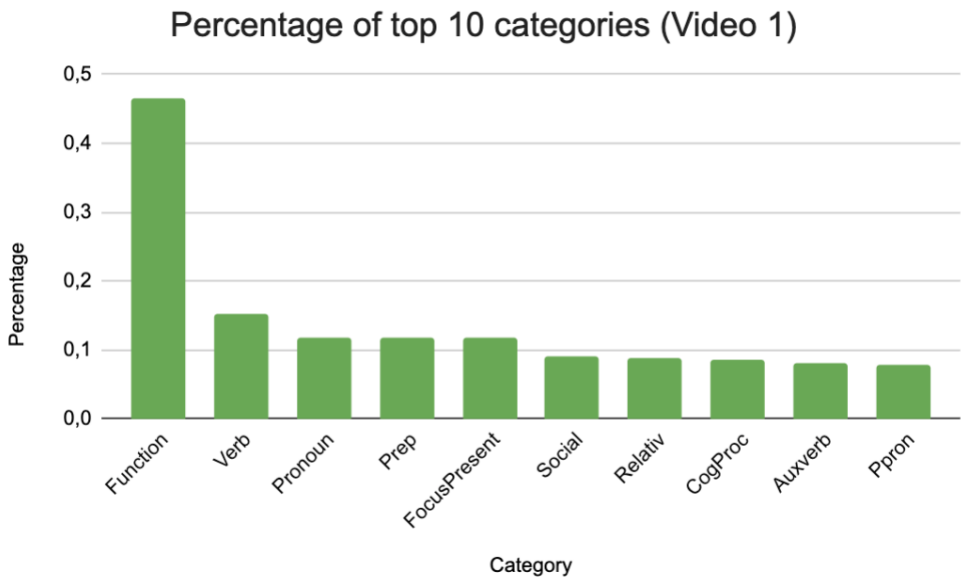
After receiving the assignments from Dr. Trung, we immediately set the first meeting on 25/04/2022 to set a detailed tentative timeline for each task. The problem given by Mr. Trung is not a simple question, it is a big problem that we have to make questions to fully understand. We broke down the big problem into smaller ones to solve task by task and easy to manage. Most importantly, we discuss together and make sure that each member understands the project guideline with the help of Dr. Trung. We had booked some meetings with him to clarify and ask questions in order clearly understand and work well.

The final project has two main parts to complete: coding and report. Therefore, we split into two teams to carry on these tasks. In the coding part, Hong Ha and Ngoc Ha are in charge of discussing and coding to provide the result for the next steps, which are analyzing the results. Because there are two videos that we have to work on, we divided into two teams to watch and annotate the data: video 1 (Hong Ha, Dang) and video 2 (Ngoc Ha, Loc) which was due 05/05/2022. Then we have a meeting on 08/05/2022 to discuss and understand the next steps, which is analyzing the descriptive statistics. Hong Ha and Ngoc Ha continued working on coding to discuss and ask questions with Mr.Trung on 13/05/2022 to finally finalize the coding part on 19/05/2022. The deadline for finalizing the report is 24/05/2022 which all of us are in charge of. To finish all the tasks in this final project, we have three meetings with Dr. Trung in total.

Results of Video 1

Scripts features

First, by using the LIWC dictionary (you can find the details here) to analyze the scripts of the video, we can make some evaluation of how people use words in daily life to describe things around them. These types of words also portray their personalities, perceptions, beliefs, etc., ... Therefore, we can have rich information about the image of street food in Vietnam in foreign visitors' eyes. There are many types of words in the provided scripts, and the main features are captured as provided in the following chart:



The top ten popular categories are Function, Verb, Pronoun, Prep, FocusPresent, Social Relative, CogProc, Auxverb, and Ppron. However, to have a general understanding of the scripts and due to the word limitation, we select the top 5 to analyze.

Approximately ½ of the scripts focus more on the images, emotions, and informal words instead of ideas formed by formal sentences because the Function words are words with ambiguous meaning and have grammatical function in a sentence, and specify the mood, beliefs, or attitudes of the speaker. These words include pronouns, determiners, and conjunctions. In daily life, native speakers these function words, which are in many informal situations. In this case, the highest percentage is of function words, which means when foreign visitors talk about Vietnamese street food, they talk in an informal way to create a friendly conversation between youtubers and the audiences, which increase the views for the channel, from which the youtubers gain more profits.

The **verbs** count for over 15%. By comparing with other top elements, it looks like most sentences contain one verb. In this case, the food is described on how they be made and how to try it properly. By this way, the verbs supplement the function words above, together create a diverse, comfortable, and detailed conversation. Therefore, enhance the joy of audience when watching this video.

The top three belongs to **pronoun** words, which count for approximately 12%. This number indicates that the youtuber review food in their own feeling. In other words, this real feeling can help audience enjoy the video to the fullest and enhance the watching experience, foster a sense of belonging to the Vietnamese street food culture.

Finally, we can see that **prep** and **FocusPresent** words ranked fourth and fifth in the chart, with exact percentage of appearance. This means many sentences contains one words representing for present and preposition. They talk right at the time they try the food, with connected words in the sentences, which create the connection between the food and the feeling of the audience and make them feel curious.

In conclusion, the friendly and attractive speech made my function words, the vivid description on the food and feelings by using **verbs** and **pronoun**, and interesting announcement made by the **FocusPresent** and **prep** words, altogether create the attractiveness of the video.

Visual Features

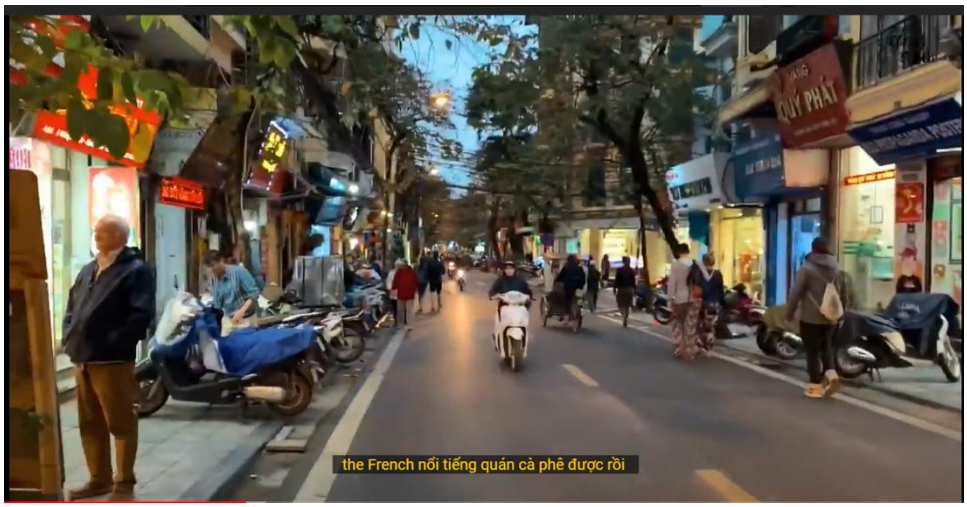
We have the following distribution:

Number of scenes	54
The total of length of the video	14 minutes 47 seconds
Average time per scene	15.407407407407407
Percentage of food existing overall	0.42592592592592593
Percentage of Youtuber existing overall	0.3333333333333333
Percentage of existing places overall	0.38888888888888889

The annotation gave us the percentage of food and Youtuber and existing places accounts for about one-third to 40% of the entire clip. This slower result on two elements made us feel curious and amazing because the video's content is about street food with the Youtuber. It can be referred that although the percentage of Youtuber and places doesn't high as we expected, it still attracts a huge number of audiences. So, it means we do not need too much images-focusing on the content to express it, just need enough.

In order to analyze the visual features of this video even further, we want to analyze the data extracted from computer visions that we have calculated above. Here's the result of that calculation.

By looking at the result of objects table, the objects that appeared the most in the top 1 object is "Gila monster", which appeared 10 times. Combined with many strange elements in the objects table like sombrero, or shield, buckler, we witness that most of the objects by AI seems incorrect.



	Start_time (seconds)	end_time (seconds)	duration (seconds)	food_focus_existing (1 or 0)	description	max_1	max_1_value	max_2	max_2_value	max_3	max_3_value
0	0	7	7	0	the scene of the street and street vendor	isopod	0,09037979404	digital clock	0,03300294658	rock python, rock snake, Py	0,04355242336
1	8	9	1	0	Scene of a restaurant on the street	slide rule, slipstic	0,4204089447	barbell	0,1224326945	maze, labyrinth	0,0973450414
2	10	12	2	1	The scene of a woman makes "Bánh mì"	slide rule, slipstic	0,3523629366	barbell	0,1365910664	maze, labyrinth	0,2379567577
3	13	18	5	1	A woman makes "Bún" on the street.	shield, buckler	0,6070776018	maze, labyrinth	0,0401115181	walking stick, walkingstick,	0,03675067466
4	19	21	2	1	Scene of a fruit cart	pole	0,3400181356	sombrero	0,1729650723	iron, smoothing iron	0,1741178625
5	23	27	4	0	Scene of cars and motorbikes running on the road	sombrero	1,238431614	velvet	0,1031560177	albatross, mollymawk	0,01518849628
6	28	29	6	0	A woman sells food for guest on the street.	theater curtain, t	0,2466699048	pelican	0,1728633092	isopod	0,04410090504
7	30	34	4	1	Scene of a sticky rice cake	theater curtain, t	0,06568656804	mosque	0,09443806758	American chameleon, anole	0,05587788273
8	35	42	7	1	Scene of egg coffee cup at Giang coffee	bakery, bakeshop	0,2444588274	mushroom	0,0917177977	triumphal arch	0,05652609881
9	43	47	4	0	Scene of the street	chest	0,2607798767	coral fungus	0,07383429584	running shoe	0,0696346405
10	48	51	3	1	Scene of the place that sells fried cakes.	typewriter keybo	0,4550292137	sombrero	0,2533905134	bath towel	0,04424603708
11	52	55	3	0	Scene of the street with the apperance of restaurants	pole	0,3793522819	iron, smoothing iron	0,1121347197	pickelhaube	0,1992203163
12	56	60	4	0	Scene of the street with the apperance of restaurants	Gila monster, He	0,5109545054	velvet	0,852978535		0,08182114561
13	61	63	2	1	Scene of egg coffee cup at Giang coffee	zucchini, courget	0,1288911611	Gila monster, Heloderma s	0,1501139041	ski	0,05509978393
14	64	73	9	0	The youtuber's legs move and the video's intro text t	sombrero	0,9620026713	pickelhaube	0,6961394979	alligator lizard	0,08006230445
15	74	106	32	0	Youtuber introduces the video and invite friends to st	plate rack	0,5156061562	zucchini, courgette	1,051058763	spiny lobster, langouste, roc	0,1248405023
16	107	109	2	1	Scene of a fruit cart	Gila monster, He	0,1565377763	lumbermill, sawmill	0,146485798	chest	0,02259195904
17	110	113	3	0	Scene of a temple in Hanoi	knot	0,1364868975	fire screen, fireguard	0,06053781302	isopod	0,05254334778
18	114	135	21	0	Youtuber introduces Hang Bac street's history and th	racket, racquet	1,740829277	pop bottle, soda bottle	0,3564305789	chest	0,5482334747
19	136	140	4	0	Scene of xoi Yen restaurant on the street	space heater	1,251155527	face powder	0,1098883672	African crocodile, Nile croc	0,03860184461

1. Image represents a part of the objects table and the scene of the Youtuber

In contrast, the color data seems accurate for most of the scenes in the video. For example, from screening the result of the colors table, the colors that appeared the most appeared the most in the are “LightSalmon4”. By observarion, the main character skin colour resembles this color. This refers to the accuracy of the color from AI.

2. Image represents the scene of the Youtuber, which appeared nearly 50% of the video

By overview the result of the places table, the places which appeared the most is beauty_salon, it's obvious that this is incorrect when using AI to calculate because almost all the locations of video was filmed at Hanoi Old Quarter on the street which consists of many vendors. Additionally, looking at the other results, we see most of the places extracted by AI are strange.

	Start_time (sec	time (sec	ction (sec	exi	xis	ce_existi	description	max_1	max_1_value	max_2	ax_2_valu	max_3	max_3_value
0	0	7	7	0	0	0	the scene of the street and street vendor	beauty_salon	1,737140004	television_studio	0,085529	slum	0,2183199937
1	8	9	1	0	0	1	Scene of a restaurant on the street	booth/indoor	0,2656599865	jewelry_shop	0,126560	nursery	0,04720000103
2	10	12	2	1	0	1	The scene of a woman makes "Bánh mì"	booth/indoor	0,2680800122	jewelry_shop	0,134540	nursery	0,05965000035
3	13	18	5	1	0	1	A woman makes "Bún" on the street.	candy_store	0,2225599933	kindergarden_classroom	0,070439	fastfood_restaurant	0,06722000139
4	19	21	2	1	0	0	Scene of a fruit cart	ice_cream_parlor	0,4381000102	bowling_alley	0,157389	bakery/shop	0,2011599961
5	23	27	4	0	0	0	Scene of cars and motorbikes running on the	ice_cream_parlor	1,823710006	chemistry_lab	0,325180	coffee_shop	0,09291000105
6	28	29	6	0	0	1	A woman sells food for guest on the street.	beauty_salon	0,4139199927	jail_cell	0,212519	elevator_lobby	0,05312000075
7	30	34	4	1	0	0	Scene of a sticky rice cake	beauty_salon	1,366319979	gas_station	0,043259	bookstore	0,033839999
8	35	42	7	1	0	1	Scene of egg coffee cup at Giang coffee	crosswalk	3,038219994	street	0,884980	bazaar/outdoor	0,008459999855
9	43	47	4	0	0	0	Scene of the street	fastfood_restaurant	0,286569997	booth/indoor	0,088750	ice_cream_parlor	0,1142100017
10	48	51	3	1	0	1	Scene of the place that sells fried cakes.	operating_room	0,193749994	beauty_salon	0,471760	art_school	0,1467999974
11	52	55	3	0	0	0	Scene of the street with the apperance of res	restaurant_kitchen	0,313909999	sushi_bar	0,176280	fastfood_restaurant	0,1842099943
12	56	60	4	0	0	0	Scene of the street with the apperance of res	beauty_salon	1,276330015	dressing_room	0,180709	shower	0,1164200004
13	61	63	2	1	0	1	Scene of egg coffee cup at Giang coffee	beauty_salon	0,9094500134	dressing_room	0,119289	shower	0,05398000032
14	64	73	9	0	1	0	The youtuber's legs move and the video's int	ice_cream_parlor	3,113100041	bakery/shop	0,392069	delicatessen	0,4057400007
15	74	106	32	0	1	0	Youtuber introduces the video and invite frie	beauty_salon	8,93882002	dressing_room	2,261150	shower	0,5762899938
16	107	109	2	1	0	0	Scene of a fruit cart	legislative_chamber	0,3326100072	classroom	0,314699	conference_center	0,1158399956
17	110	113	3	0	0	0	Scene of a temple in Hanoi	street	0,4874100007	arcade	0,064159	bazaar/outdoor	0,0595900009
18	114	135	21	0	1	0	Youtuber introduces Hang Bac street's history	fastfood_restaurant	0,5144300021	food_court	0,231770	beauty_salon	5,130939972
19	136	140	4	0	0	1	Scene of xoi Yen restaurant on the street	beauty_salon	1,861090004	dressing_room	0,331370	locker_room	0,110020001
20	141	157	16	0	1	0	Youtuber invites people to come a famous pla	beauty_salon	2,520660044	gas_station	0,332820	hardware_store	0,2085199973
21	158	167	9	0	0	1	Introduces what is in Xoi Yen	ice_cream_parlor	1,462010027	bakery/shop	0,467499	delicatessen	0,3398000104
22	168	178	10	1	0	1	Xoi Yen's staff brings Xoi to the youtuber's tal	beauty_salon	1,00258001	dressing_room	0,435590	classroom	0,3215400058
23	179	184	5	1	0	1	Close-up of the plate of sticky rice and other	beauty_salon	0,5512499962	ice_cream_parlor	0,394640	restaurant_kitchen	0,5258900054
24	185	225	40	1	0	0	Take a close-up shot of each dish on the table	street	1,000969986	crosswalk	0,169540	residential_neighborhood	0,08534999934

3. Image represents a part of the places table

By observation from the scene table, the scenes that appeared the most is the human scenes named “person;individual;someone;somebody;mortal;soul”. While comparing this scene and other scenes, we see that most of the scenes are correct.

	Start_time (sec	time (secc	tion (seco	us_existin	existi	ce_existi	description	max_1	max_1_value	max_2	max_2_value	max_3	max_3_value
0	0	7	7	0	0	0	the scene of the street and stre	person;individual;some	2,527160645	road;route	0,4307522244	wall	1,141859266
1	8	9	1	0	0	1	Scene of a restaurant on the str	wall	0,3426920573	signboard;sign	0,2399359809	base;pedestal;stan	0,234375
2	10	12	2	1	0	1	The scene of a woman makes "B	wall	0,4270155165	base;pedestal;stand	0,2398071289	signboard;sign	0,1621839735
3	13	18	5	1	0	1	A woman makes "Bún" on the s	arcade;machine	0,8712836372	table	0,5497707791	wall	0,08635796441
4	19	21	2	1	0	0	Scene of a fruit cart	plate	1,303697374	table	0,4843139648	person;individual;s	0,6561550564
5	23	27	4	0	0	0	Scene of cars and motorbikes ru	plate	2,377977159	glass;drinking;glass	0,1003011068	table	0,1975165473
6	28	29	6	0	0	1	A woman sells food for guest or	person;individual;some	1,128370497	wall	0,3408881293	floor;flooring	0,2295396593
7	30	34	4	1	0	0	Scene of a sticky rice cake	building;edifice	1,117180718	person;individual;someon	2,031901042	poster;posting;plac	0,004272460938
8	35	42	7	1	0	1	Scene of egg coffee cup at Gian	person;individual;some	2,471896701	road;route	1,299065484	building;edifice	0,5030992296
9	43	47	4	0	0	0	Scene of the street	wall	2,139153375	ceiling	0,3741319444	refrigerator;icebox	0,04988606771
10	48	51	3	1	0	1	Scene of the place that sells frie	person;individual;some	1,927062988	table	0,6213650174	wall	1,064812554
11	52	55	3	0	0	0	Scene of the street with the app	table	0,5796101888	person;individual;someon	1,531494141	tray	0,5204806858
12	56	60	4	0	0	0	Scene of the street with the app	person;individual;some	1,728820801	wall	1,193020291	basket;handbasket	0,03110080295
13	61	63	2	1	0	1	Scene of egg coffee cup at Gian	person;individual;some	1,295756022	wall	0,6590711806	tray	0,08957926432
14	64	73	9	0	1	0	The youtuber's legs move and t	table	1,486456977	tray	1,515333388	plate	0,7521701389
15	74	106	32	0	1	0	Youtuber introduces the video a	person;individual;some	14,17437744	wall	12,44869656	plate	0,1693386502
16	107	109	2	1	0	0	Scene of a fruit cart	person;individual;some	1,519354926	wall	0,5733303494	flag	0
17	110	113	3	0	0	0	Scene of a temple in Hanoi	road;route	0,3482801649	building;edifice	1,006591797	person;individual;s	1,766215007
18	114	135	21	0	1	0	Youtuber introduces Hang Bac s	wall	5,379489475	ceiling	0,2625732422	counter	0,2763400608
19	136	140	4	0	0	1	Scene of xoi Yen restaurant on t	person;individual;some	3,287801107	wall	1,346476237	building;edifice	0,1837497287
20	141	157	16	0	1	0	Youtuber invites people to com	person;individual;some	6,250868056	wall	3,889960395	building;edifice	0,5895453559
21	158	167	9	0	0	1	Introduces what is in Xoi Yen	tray	0,5936821832	food;solid;food	0,5461968316	wall	2,263000488
22	168	178	10	1	0	1	Xoi Yen's staff brings Xoi to the	person;individual;some	4,439066569	wall	2,625128852	flag	0
23	179	184	5	1	0	1	Close-up of the plate of sticky ri	person;individual;some	2,855719672	windowpane;window	0,324605306	wall	1,03982883
24	185	225	40	1	0	0	Take a close-up shot of each dis	tray	0,7615288638	minibike;motorbike	0,7276068703	building;edifice	3,600504557

4. Image represents a part of the scene table and the scene of the Youtuber

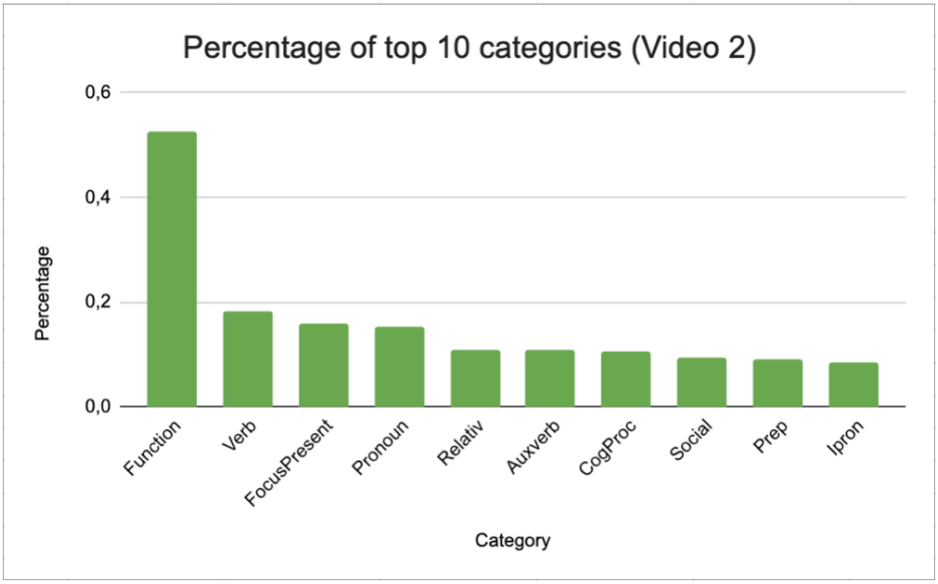
To conclude, combined with other observations, objects detected by the machine are foreign and incorrect in which video took place. In combination with the objects, the places recognized the most were not the location of the filming. However, the color seems to be easier to distinguish and quite accurate according to what we observed and most of the scenes are also accurate. From this analysis, we also knew that the appearance of the Youtubers in most of the scenes can attract the attention of the audiences. But it doesn’t need too much.

To make a discussion, from our perspectives, the reasons for these mistakes can be rooted in the lack of data, or human biases while compiling the data. But one bright side we can take

from this video is that computers support us a lot more in identifying objects we often miss while watching. These results suggest us to train more data in the field of Vietnamese food objects and Vietnamese places for this AI model to achieve more accurate result. It also provides an advice for Youtuber and communicator in developing scripts and visual data of spreading Vietnamese street food.

Results of Video 2

Script features



After taking top categories of words like the previous video, we reached this chart which represents top 10 categories of words in video 2. Looking at the chart, we can see that like the previous video, approximately 50% of the scripts focus more on **function words,** which means the video focus on the images, emotions, and informal words instead of ideas formed by formal sentences of the scripts. This makes the audience feel comfortable and follow the channel with daily vibes. And this can increase profits and viewers for YouTubers.

While the **verbs** count for nearly 20%. We can say that because all sentences mostly contain verbs, so it just a normal case. But we can conclude that the existence of verbs adds to the creation of vivid function words, helping the diverse and comfortable conversation in the video.

Moreover, we can see that **'FocusPresent'** ranked third in the chart. This means most of the sentences contains one words representing for present. In other words, they mostly talk about the present. Talking about the present made the audience feel lively and energetic about the scene. The using of words here also refers to the changing of positions and actions, which avoid

the boring moment with repetition and move the audiences to many places with the announcement. These changes will make the audience focus and increase their curiosity, then they will increase the time watching in each video. In that way, it will also attract more audiences and increase the profits of the Youtuber.

In the top five, the use of **relati** words also means that they talked much about the relations, maybe they are traveling to many places. This obviously means they talk too much with “relativ” words. Moreover, combining with the use of **social words**, counting for nearly 10%. This interprets that the Youtuber always talked about the relationships between humans and objects. This will make the audience feel the real connection between them and the Youtuber, the people, and the things in the video. Then, they will find these video interesting.

To conclude, the combination of (1) attractive, comfortable language made by function words, (2) the diverse changes and relations made by "relative, prep", and (3) the interesting announcement made by "focus present" will attract the audiences, make them feel curious, and highten the number of followers.

Visual features

When we run the annotated file to calculate its information, we received the information of the annotated file in the table below:

Number of scenes	58
The total of length of the video	14 minutes 42 seconds
Average time per scene	14.224137931034482
Percentage of food existing overall	0.46551724137931033
Percentage of Youtuber existing overall	0.6379310344827587

From the information of the annotated file, we can see the high percentage of the food existing and Youtuber existing. This refers to the attraction of the main content, food, and the Youtuber, will combine to make more and longer audiences. Then, we will use these number and the annotated file to analyze in the later parts with focusing on 4 elements, namely color, scenes, objects, and places.

The colors analyzed by the computer are **mostly correct**. For example, the couple in the video both wearing grey clothes, and the grey appears most of the time when the YouTuber appears in the video. AI data seems the same as with the annotated scene).

gray16	7565	gray20	6619	gray17
gray100	26465	LightSlateGray	7737	maroon
SkyBlue4	62299	LightSkyBlue4	59290	burlywood4
gray100	26959	gray7	21869	gray6
gray70	4771	gray69	4217	gray100
LightSalmon4	23416	gray100	29913	gray20
gray8	94024	gray9	118868	gray7
gray64	22866	gray63	17121	gray65
gray64	47765	gray63	49681	gray6
gray65	8701	gray69	8042	gray63
gray0	493126	gray1	48719	gray2
gray100	12320	GhostWhite	8857	AliceBlue
gray14	33974	gray15	29794	gray13
gray100	37912	gray99	10936	gray60

1. Image represents a part of the color table and the scene of the Youtuber

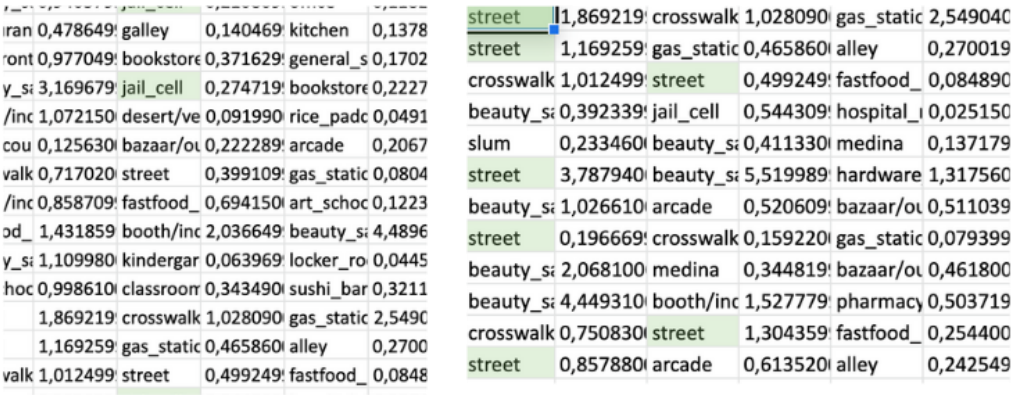
Looking at the objects table, the objects analyzed by the AI seem totally incorrect and strange because many object analyzed by the computer came from nowhere. Most of the food focusing scene, created by annotation, was identified with a strange object, which is not food. For example, Ciconia nigra or Spoonbill have nothing related to the video. Or most of the food focusing scene was identified with a strange object, which is not food (same objects table below)

reflex can	0,194964		0,378332	reflex can	0,197206	pole	2,159266	chest	0,425396	pole	0,490619
isopod	0,158582	pop bottl	0,050784	isopod	0,043149	lacewing,	0,070625	leatherba	0,063767	lacewing,	0,022357
great grey	0,405143	American	0,324386	great grey	0,119824	pole	0,529978	chest	0,186859	pole	0,129262
chest	2,314523	pole	4,438098	chest	0,485102	zucchini,	0,320495	jersey, T-s	0,041362	zucchini,	0,042835
racket, ra	0,358609	axolotl, m	0,090618	racket, ra	1,429866	chest	0,814938	American	0,125512	chest	0,117933
Crock Pot	0,872938	waffle iro	0,307459	Crock Pot	0,042125	chest	0,459420	pole	0,411309	chest	0,280377
racket, ra	0,190068	American	0,096092	racket, ra	0,087616	horizonta	0,299808	pole	1,172249	horizonta	0,117974

1	The scene of a woman makes "Bánh mì"	slide rule, slipstick
1	Scene of a fruit cart	Gila monster, Heloderma suspectum
1	Xoi Yen's staff brings Xoi to the youtuber's table	great grey owl, great gray owl, Strix nebulosa

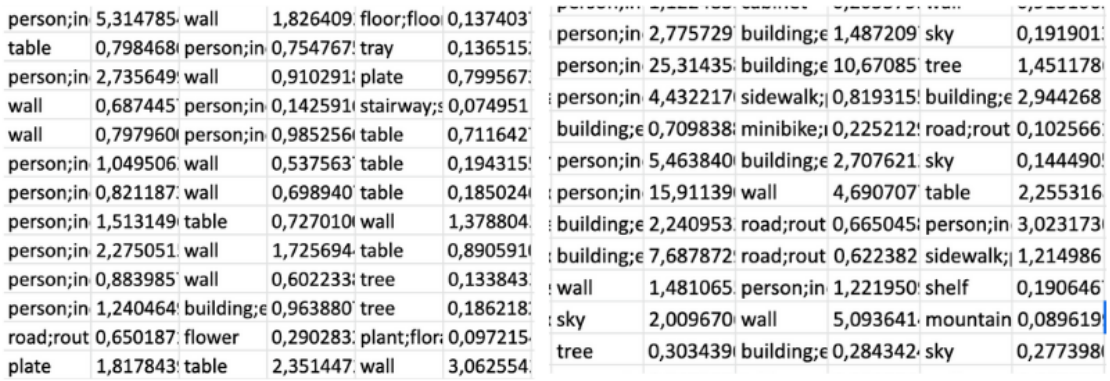
2. Image represents some cells in the objects table and the scene containing food object

Many places which were analyzed by AI are correct. However, there are some unreasonable places like Jail_cell. Or there are many eating places that were identified wrong by the AI to be beauty_salon, martial_arts_gym,...



3. Some wrong identified eating places in places table

In the scene table below, Scenes named "person;individual;someone;somebody;.." appeared the most for 57% of the annotated scene. The number of Youtuber who appeared in annotated scene file is 46%. This number reflects the accuracy of AI in realizing humans. Moreover, we also saw that most of the scenes analyzed by the AI are mostly correct.



4. Some cells in the scene table

To conclude, AI can identify each pixel in color easily and the scenes' data set doesn't different in many different urban areas. The big percentage of humans represents for the attraction of the scene with humans in a food-oriented video. However, AI's data set isn't diverse enough to identify all objects in Vietnam, especially Vietnamese food. The places are more diverse but still limited which led to many wrong places. While AI has been developed a lot, it still contains many limitations.

Conclusion

We knew that the scripts feature analysis let us know some word tips to attract more audience with longer watching time and the visual feature analysis let us know about the visual features that built a video Youtube on street food and attracts more audiences, especially with humans and food scenes. From that, the comparison between AI and annotated let us know the difference between data extracted by annotated and AI. And we know the AI's limitations, raise awareness, and improve it in the future.

At the beginning, we had no idea that computer science covered such a wide variety of topics. It is not enough to know how to write code and create programs; we must also understand the world around us and be able to apply what we learn to the development of innovative technologies. We also gained an understanding of the significant influence that computer science has on our day-to-day lives. It can be found, for instance, in films, songs, video games, and even the food that we eat.

When it comes to studying computer science, there are a lot of areas in which we can make improvements. First and foremost, we need to have a solid foundation in fundamental aspects of computer science. By gaining an understanding of the various concepts and how they are related to one another, we can acquire the foundational knowledge necessary for studying computer science. The second thing is that we need to be able to think critically and creatively at the same time. This will assist us in overcoming challenging problems, coming up with fresh ideas, and developing original approaches to solving problems. The third requirement is that we should be capable of working together effectively and having clear and concise communication. This will be helpful for us when working on a project as part of a team or when discussing our ideas with others who might have a different point of view than us.

Last but not least, we suggest Loc should improve his coding skills. Moreover, Khanh Dang should improve her time management to seriously work on the final project and learn more about Literature review writing and details analysis, which would truly contribute to our project instead of wasting our time and being irresponsible.

References

Hiemstra, A. M., Van der Kooy, K. G., & Frese, M. (2006). *Entrepreneurship in the street food sector of Vietnam—Assessment of psychological success and failure factors*. Journal of Small Business Management, 44(3), 474-481.

Hoang, D. S., & Tučková, Z. (2021). The impact of sensory marketing on street food for the return of international visitors: A case study in Vietnam. Scientific Papers of the University of Pardubice, Series D: Faculty of Economics and Administration.

Linh, P. L. D. (2021). *TOURISTS' EMOTIONAL RESPONSES TO STREET FOOD EXPERIENCES IN VIETNAM (Doctoral dissertation, University of Surrey).*

Samapundo, S., Thanh, T. C., Xhaferi, R., & Devlieghere, F. (2016). Food safety knowledge, attitudes, and practices of street food vendors and consumers in Ho Chi Minh City, Vietnam. *Food Control, 70, 79-89.*

Stutter, Natalia 2017. *The social life of street food: exploring the social sustainability of street food in Hanoi, Vietnam*. PhD Thesis, Cardiff University.

Wikimedia Foundation. (2022, May 30). *Function word*. Wikipedia. Retrieved May 31, 2022, from https://en.wikipedia.org/wiki/Function_word#:~:text=In%20linguistics%2C%20function%20words%20(also,or%20mood%20of%20the%20speaker.