VizWiz-VQA disaggregation: A new set of visual question classes.

👨‍🔬 **Hernán J. Maina**[1,2] and 🕵️‍♀️ **Laura Alonso Alemany**[1,2]

[1] Universidad Nacional de Córdoba
[2] CONICET, Argentina
hernan.maina@mi.unc.edu.ar, alemany@unc.edu.ar

Abstract. *VizWiz-VQA (VizWiz Visual Question Answering), is the first data-set of visual questions answers made for and by blind people. It, structures his questions into four main categories, where 'other' is the predominant class with around 65%, followed by a salient set of questions (∼27%) classified as 'unanswerable'. In this work, through the exploration of different clustering strategies and morpho-syntactic analysis, a new set of eight main categories is presented and proposed, which are used to fine-tune a automatic classification model and in this way to be able to reanalyze the original set from a new perspective. It should be noted that this research sets aside the visual modality 'V', to focus on the 'QA' part of VQA, with the aim of disaggregate the majority classes, to facilitate the understanding of the nature of the questions, and the reasons for why many of these questions cannot be answered.*

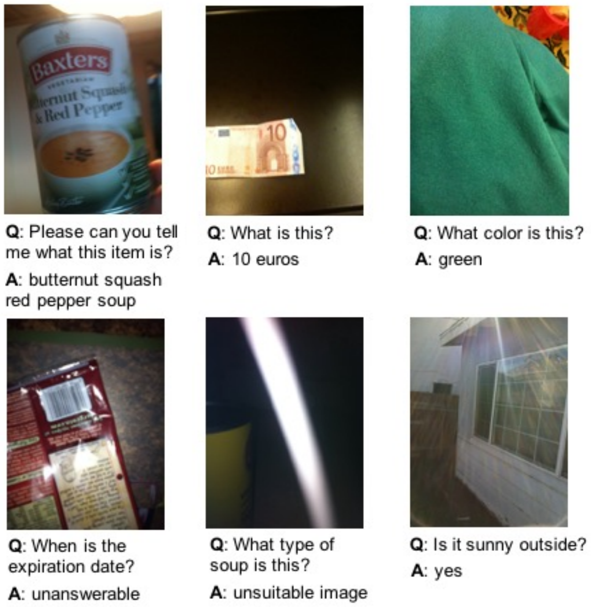
Q: Please can you tell me what this item is?
A: butternut squash red pepper soup
Q: What is this?
A: 10 euros
Q: What color is this?
A: green
Q: When is the expiration date?
A: unanswerable
Q: What type of soup is this?
A: unsuitable image
Q: Is it sunny outside?
A: yes

Fig. 1: Samples of VizWiz-VQA dataset.

1 Introduction

VizWiz[3], is the project responsible for introducing the first datasets and challenges, aimed at motivating the creation of new and better AI assistive technology/algorithms, to help people with visual impairment. In 2010, Bigham et al. [2] presents VizWiz, a mobile phone application designed to help blind people with their daily problems. Users took a photo and recorded a question about what they wanted to know about that image. Such visual questions were answered by a group of workers, mostly recruited from companies like Amazon Mechanical Turk. Eight years later, Gurari et al. [4] filtered and anonymized the information from all the data collected up to that point, and for each pair of images/questions, they collected a total of 10 answers using crowd-sources. In the Computer Vision and Pattern Recognition 2018 conference (CVPR 18), they presenting the first non-artificial 'goal-oriented' public VQA data-set, built entirely from data originated by blind people. Figure (1) illustrates some examples of this dataset.

VizWiz-VQA. Since the same blind people were responsible for capturing the photographs and recording the questions, a large part of the images in this data set are characterized by focusing, lighting and framing problems. On the other hand, since people speak differently from how they write, the data set is markedly conversational. With a large majority of incomplete questions due to cuts and imperfections in the audios from which they were transcribed. As a result, either because the question could not be answered from the context of the image, or because the image quality was unsuitable, VizWiz-VQA has a large number of visual questions that cannot be answered, unlike other VQA datasets.

Specifically, VizWIz-VQA[4] is conformed of 20523 images/questions pairs, and 205230 associated answers in the training set; 4319 image/question pairs, and 43190 associated answers in the validation set; and 8000 pairs of images/questions in the test set. Each of the questions is assigned a category, which

[3] https://vizwiz.org

[4] https://vizwiz.org/tasks-and-datasets/vqa/

inherits the type of expected answer (the values in parentheses represent the frequency in the dataset): ‘number’ (∼1,4%), ‘yes/no’ (∼4.63%), ‘unanswerable’ (∼27.84%) and ‘other’ (∼66.1%). The average length of each question is around ∼6.8 words, and approximately 28% of the opening words appear less than 5% of the time in the set. The latter, attributed to the conversational nature of the set, where questions such as “Hi, can you please tell me which is . . . ” in VizWiz, it is found simply as “Which is. . . ” in datasets like VQA v2.0[5].

Contributions. In this work, as a consequence of the analysis of ∼24.800 pairs of questions and answers from the training and validation sets, the following contributions were made:

– Exploration and identification of pre-processing and clustering strategies, for the consistent disaggregation of groups of majority questions of the VizWiz-VQA dataset.
– Definition of eight new main categories, to identify and characterize in a more descriptive and natural way, the total set of VizWiz-VQA questions.
– Training and testing of the automatic classification model on the new defined categories, and analysis of the original categories, mainly the ‘other’ and ‘unanswerable’ clases, based on the new classifications obtained.

2 Related works

In 2013, Brady et al. [3], launches VizWiz-Social, an update of the application proposed three years ago by Bigham et al. [2], which they use for a long year with the aim of providing a new look at the diversity of questions that blind people want answers about their visual environment . This was one of the first works to perform a qualitative analysis to build a taxonomy of the types of questions asked of blind people. Although the classification process was not performed using unsupervised learning algorithms, it laid the groundwork for improving understanding of the problems faced by blind people in their daily lives.

While there are no papers that specifically disaggregate the main categories of the VizWiz-VQA dataset automatically, many clustering approaches and strategies have been used to group questions into other datasets for a particular task.

Aishwarya Ashok et al. [1], used clustering on OQA (Opinion-based Question Answering) datasets, to answer questions about online stores, based on the opinions left by other customers, using cosine similarity scores between revision and question sentence vectors.

Kento Terao et al. [7], proposes a novel approach to identify the level of difficulty in the visual questions of VQA (specifically about VQA 2.0), grouping the levels of difficulties in relation to the entropy values calculated based on the distribution of the responses to each question.

Lastly, Deepak P. [6], applies clustering on CQA (Community-based Question Answering) questions, in which it uses a novel strategy to group question-answer pairs in data sets collected from systems such as Yahoo! Answers, Stack Overflow, etc.

3 Clustering

Mainly, the question disaggregation process of the VizWiz-VQA dataset was carried out using as a backbone, the classic unsupervised learning algorithm called *KMeans* from the sklearn library[6]. This, as it does not require supervision or prior labeling, is very useful for exploring datasets with unknown structures and distributions. Through the iterative determination of centroids, this method allows data with similar characteristics to be grouped according to their distance from the closest centroid.

Like any other machine learning model, KMeans also requires that its input data be numeric representations. For this reason, various strategies were tested to map sequences of words (questions, questions + answers, etc.) to their respective feature vectors.

3.1 Data preprocessing

As a first phase, a curing process of the VizWiz-VQA dataset was carried out. As all the analyzes were carried out on the questions and the answers, only the data corresponding to the training and validation sets were used, since the test set does not contain associated answers.

The first step was to normalize each of the questions. Using the *contractions*[7] library, the contractions were expanded, non-alphabetic characters filtered out, and converted to lowercase. Then, after deleting the leading and trailing blanks, with the help of the NLP *spaCy*[8] library, the questions were filtered with more than one sentence. Later, the duplicate questions were also eliminated. Remember that as this is a conversational dataset, many questions are very long and contain irrelevant information that could introduce noise into the clustering process.

The result of these procedures culminated in 9003 unique pairs of questions and answers, out of an initial total of 24842 pairs. In the process, the 10 responses

[5] `https://visualqa.org/`

[6] `https://scikit-learn.org/stable/index.html`

[7] `https://pypi.org/project/contractions/`

[8] `https://spacy.io/`

associated with each question were re-ordered in descending order in a list of tuples of the form (response, number of matches). Figure (2), shows the final distribution of the 50 most frequent words contained in the 9003 questions. On the other hand, in Figure (3), the 100 most frequent answers are displayed as cloud word style, considering (not considering) the two most frequent answers 'unsuitable' and 'unanswerable' respectively.

3.2 Data representations

For the representation of the data, two different types of embedding were used: I) embedding based on occurrence matrices + dimensionality reduction; and II) embedding obtained from pre-trained neural models.

For type (I), different combinations of input data concatenations were tested. Using the *spaCy* library, the following were tested: 'question lemma list', 'question lemma lists + best answer lemma list', 'question lemma list + all answer lemma list'; in the same way with combinations of words without stemming and PoS (Part of Speech) of the questions. It should be noted that in the resulting concatenations, in order to differentiate the lemmas belonging to the questions from those belonging to the answers, the special tokens **CLS** and **SEP** were added, located at the beginning and end of each list.

The construction of the occurrence matrices were performed using different ranges of n-grams, from (1-gram,..,3-gram) to a maximum of (1-gram,...,10-grams). After this, the dimensionality reduction was performed using a variance threshold. It was implemented through the *varianceThreshold* method of the *sklearn* library, providing vectors with dimensions that ranged from 90 to 120.

For the type of embedding (II), encodings were tested using pre-trained models at the word level, using *fastText*, and at the sentence level, using *doc2Vec*; both from the *gensim*[9] library. For the particular case of the *fastText* model, two strategies were used to obtain the resulting embedding: summation of individual embedding on the one hand, and multiplication of individual embedding on the other.

3.3 First approach

As a first approximation to the problem, a succession of executions of the KMeans algorithm were carried out, using different ranges of k values, where $k = number\_of\_desired\_clusters$. The evaluation of the results was carried out visually and qualitatively, with the help of the generation of Silohuette graphs that guided the selection process of the most optimal k values. In general terms, in this first contact, we sought to reduce the set of variables and configurations that would yield better groupings, and to know the background of the behavior of the results.

From such tests, it was observed that many questions like: emph 'please can you tell me what's in this box? ' And emph 'what's in this can?', They were placed in different clusters, despite having similar semantics. After conducting a detailed review of the complete set, it was identified that certain conversational sentences located at the beginning of the questions, were responsible for confusing the grouping algorithm. A complete list of all identified sequences can be found in the Annex (A). This finding forced a new data re-processing, and re-generation of lists of 'lemmas', 'PoS' and 'tokens' used in the generation of each embedding. Now, every time a statement of these characteristics was identified as a sub-sequence of a question, its tokens were not added to the lists. Another important aspect observed was that input combinations based on Part of Speech lists and non-stemmed words, gave much less accurate results than when using different combinations of lemmas lists.

3.4 Selection of the best strategy

With all the information obtained from the first approximation (3.3), and given the main objective of disaggregating the majority categories of VizWiz-VQA through the groups of questions returned by the clustering algorithm; In this second stage, more systematic testing methods were incorporated in order to quantify the quality of the results.

One of them was the well-known *Elbow* method. This is a heuristic used to determine the number of clusters in a data set. This method graphs the variability as a function of the number of clusters, and selects the 'elbow' of the curve as the number of clusters to use. Although wide ranges of values were tested (between 3 and 20), all the curves obtained had a fairly linear behavior, so their results were not very useful.

As a second method, a test dataset was made, made up of N tuples of random questions, selected from among the 9003 of the filtered data set. Each pair was labeled with Yes/No, according to whether the questions should/or not, belong to the same cluster. In total, 230 pairs were labeled, and the percentages of correct answers were calculated, both for those in which the tuple of questions fell into the same cluster, coinciding with the manual annotation, and vice versa for those who did not. Although the latter method yielded very useful data for selecting the best strategy, additional human supervision was required to determine the validity and consistency of the results. Many of them, despite indicating good Silohuette indices,

[9] `https://pypi.org/project/gensim/`

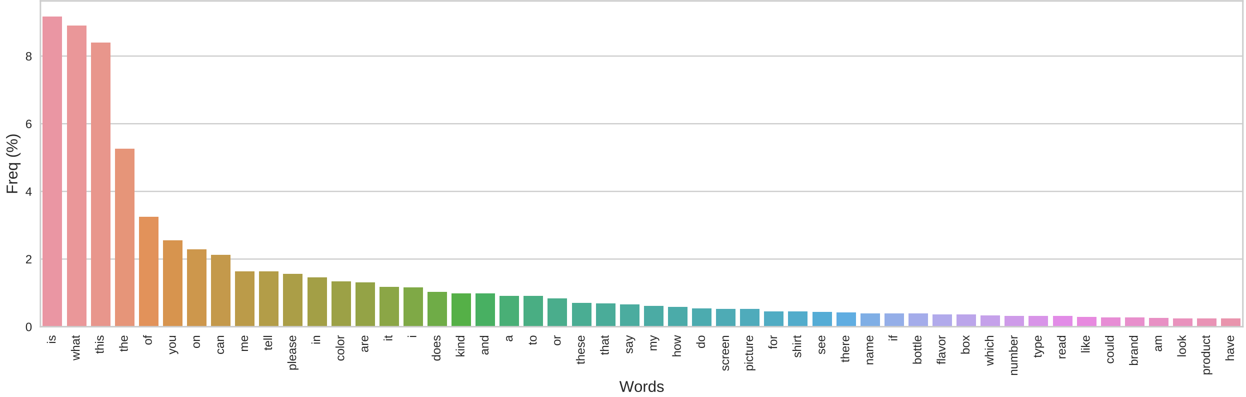

Fig. 2: The 50 most frequent words in VizWiz-VQA questions dataset.

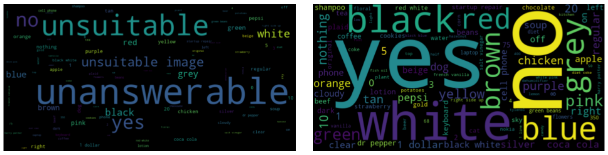

Fig. 3: The 100 most frequents answers in the VizWiz-VQA dataset. (Left): With 'unsuitable' and 'unanswerable' answers. (Right): Without 'unsuitable' and 'unanswerable' answers)

and high accuracy percentages in the test dataset, grouped questions by type and morphology, and not by semantic similarity.

In total, leaving aside the tests carried out in the first approximation, nine different clustering strategies were explored (see Table (1)). The final selection was made looking for a balance between quality/consistency of clusters delivered, and percentages of correct answers in the set of test data prepared.

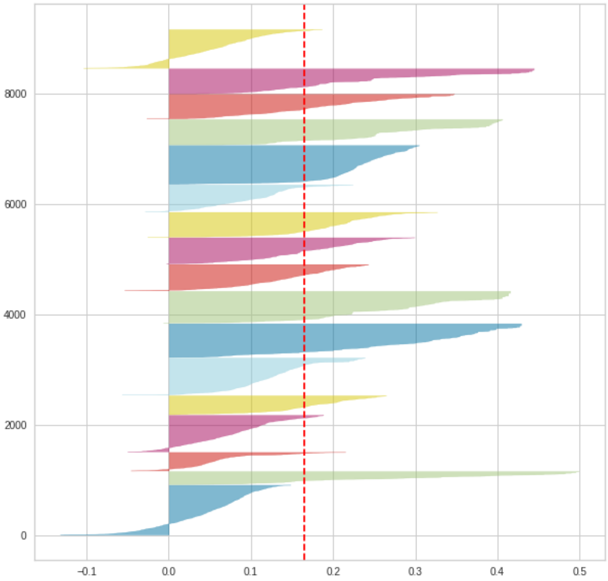

Fig. 4: Clusters consistency, through Silhouette scores.

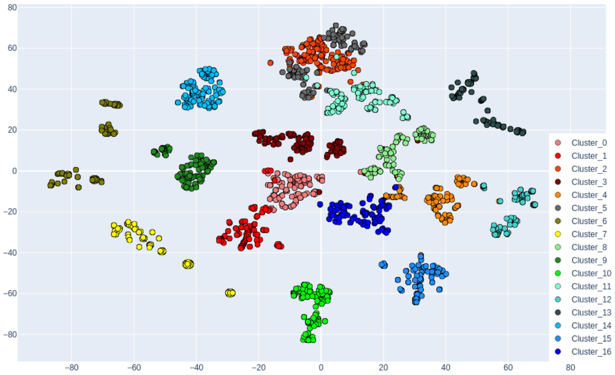

Fig. 5: Clusters distributions, displayed through T-SNE method.

Best strategy . The final strategy selected delivered **17** question clusters, see Annex (B). Figures (4) and (5), show the consistency level 'Silhouette graph', and the approximate distribution respectively. The second, after the application of T-SNE[10], to be visualized in a two-dimensional graph.[11]

Regarding the input data, these resulted from the concatenation of the lists of: **'question lemmas + best answer lemmas'**; while the embeddings that fed the Kmeans algorithm were based on an occurrence matrix with **1-grams, ..., 10-grams** as feature columns, with a subsequent dimensionality reduction, using the technique of variance threshold, which discarded columns of characteristics with variations less than **.001**, producing final vectors with a dimensionality of **99**. This strategy also gave a **44.26%** precision in the control tuples annotated as belonging to

[10] `https://en.wikipedia.org/wiki/T-distributed_stochastic_neighbor_embedding`

[11] Although the identification of the clusters in the Figure (5) corresponds to the indices of the clusters described in Annex (B), the same correspondence might not be direct in the Silohuette graph in Figure (4), since its generation was carried out separately.

Data Input	N-grams	Min_df	Var Threshold	Embedding	Best K	Same cluster	Diff cluster
Qsl	1,3-grams	10	.01	117	8	47.54%	91.41%
Qst+QsPoS+BestAns	1,10-grams	10	.001	97	19	36.07%	94.4%
Qst+BestAns	**1,10-grams**	**10**	**.001**	**99**	**17**	**44.26%**	**99.39%**
Qst+BestAns (w/ Noun mask)	1,7-grams	10	.001	113	17	47.5%	96.9%
Qst+AllAns	1,3-grams	10	.001	74	25	26.23%	96.9%
Qst+BestAns (Doc2Vec)	4-grams	1	-	100	18	14.7%	87.7%
Qst+AllAns (Doc2Vec)	4-grams	1	-	100	14	29.5%	90.8%
Qst+AllAns (FastText-vSum)	4-grams	1	-	100	21	13.1%	97.5%
Qst+AllAns (FastText-vMult)	4-grams	1	-	100	16	24.5%	87.7%

Table 1: Results and metrics of clustering strategies. (Qsl): List of question lemmas - (Qst): List of question words without lemmatize - (WsPoS): List of question's Part of speech - (BestAns): List of lemmas of answer with most agreement - (AllAns): List of lemmas of all answers from question.

similar clusters, and a **99.33%** for those annotated as belonging to different clusters (Figure (6)).

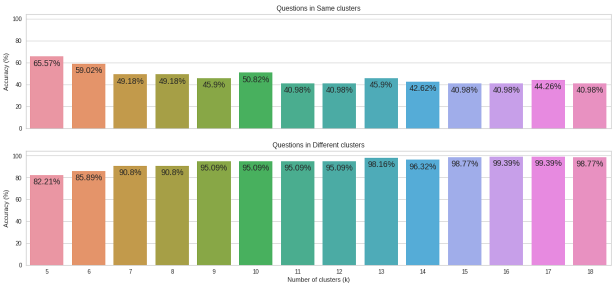

Fig. 6: Percentages of accuracy of the tuples considered to belong to the same clusters, and those considered to belong to different clusters, in relation to the results returned by the 'KMeans' algorithm using a given k value.

Based on the results and experiments carried out, it was observed that the clustering strategies based on type II embedding were much more inefficient than those based on occurrence matrices for some range of n-grams. Such observation can be attributed to the relative small amount of data used for the generation of the embedding, added to the fact that the models used were not pre-trained in datasets of a conversational nature. On the other hand, the best performances were achieved using as input data, a concatenation of the question and answer lemma lists. It was also observed that the inclusion of answers with fewer coincidences produced a negative effect. Regarding the quality of the clusters, it was noted that as the questions move away from the centroids (they gain greater distance), some consistency is lost. In general, sets of questions were obtained that share greater morphological than semantic similarity, and are moderately influenced by their answers. Despite this, clusters such as 0, 1, 5 and 15 were of great quality and purity, and combinations such as 12 and 13, 16 and 4, could be characterized very precisely.

4 Classifier

The main objective of training a classifier using the clusters of questions obtained in Subsection (3.4), was to achieve the disaggregation of the majority groups of VizWiz-VQA, by proposing new categories detailed in Subsection (4.1). Subsection (4.2), describes the training process for the different trained classification models. Later in Section (5), the most accurate models will be tested and contrasted on the total set of VizWiz-VQA data, the old categorizations, of the new classes returned by the models.

4.1 Categories assignment

Although most of the groupings found in Subsection (3.4) already had some coherence, the KMeans algorithm also performed the separation of sets based on the answers, leaving, for example, closed questions answered with 'yes' on the one hand, and answered with 'no' for the other. This is why the classification model will not be fed directly with the content of each cluster, without first performing a previous refinement step.

For the characterization and/or definition of the new classes, first the purest clusters were identified, that is, those groupings that only contained a particular type of questions, without considering their answers, for example: colors, identification of objects, questions with options, closed questions. Then, among the remaining clusters, shared characteristics were searched that could group and be representative for two, three or more of them.

As a result, **8 new classes** were defined, capable of representing in a natural way 15 of the 17 analyzed groups. Given their varied composition, the remaining clusters (7 and 9 of Annex (B)), were left aside, to avoid introducing errors in the training process of the classification model. The following items describe and exemplify the new proposed classes:

- **c0) color**. Color identification: Questions asked with the intention of obtaining information about the color of a certain object. *e.g: 'what colors is my jeans?', 'what is the color for this laptop?', 'which color has the purse?', 'what color is my t-shirt please?'*

- **c1) ocr**. Need for ocr: Questions directly aimed at obtaining specific information (textual or numerical) that helps to complete the identification of a previously identified object. e.g: *'what is the name of this film?', 'what is the expiration date of this almond milk?', 'what is the title of this disc?', 'what is the phone model?'*
- **c2) observation**. Observations: Questions where the person needs to know an appreciation or obtain textual or visual information of some characteristic of an object or scene in order to be informed. *e.g: 'what does this box say on top?', 'this is sky look like?', 'what does this computer screen say?', 'what does this pregnancy test show?'*
- **c3) ident**. Direct identification: Direct question for the identification of an object, or some property or characteristic that allows to finish identifying it. e.g: *'what is this recipe?', 'what brand of earbuds are these?', 'what kind of battery is this?', 'coffee is this?', 'what type of tile is this?'*
- **c4) rel_ident**. Relative identification: Object identification question, through referential descriptions that involve already located or known objects. e.g: *'can you see what is in this package?', 'what is on my shelves?', 'what is inside this?', 'what is written in screen?', 'what is inside this canned good?'*
- **c5) explication**. Complex answer questions: Questions with several objectives, whose formulation of the answer requires knowledge of what is being asked or involves giving location instructions, recognizing people or giving an explanation of a random topic. e.g: *'where is this made?', 'who is this dog?', 'why is this computer not booting up?', 'where is this box from?', 'who is this mail for?', 'where you thinking about this one?'*
- **c6) choice**. Choice selection: Questions where the answer is explicit in the question, and one of the listed options must be returned. e.g: *'is this iphone or nokia?', 'is this blue or purple?', 'is this decaf or regular coffee?', 'is this brown rice or white rice?'*
- **c7) yes_no**. Confirmations: Status questions. Binary response (yes/no). e.g:. *'is this the new apple keyboard?', 'are those piano keys?', 'is this the blue?', 'is this an iphone?', 'is my light off?', 'is he fat?', 'see anything?'*

4.2 Training

Before starting to train the classification models, the questions of each cluster were identified and labeled with the new proposed classes, see Table (2).

Ref	Class	Clusters assigned
c0	color	[0]
c1	ocr	[16,3]
c2	observation	[1]
c3	ident	[2,8,11,6]
c4	rel_ident	[14]
c5	explication	[4,10,15]
c6	choice	[5]
c7	yes_no	[12,13]

Table 2: Classes proposed and cluster assignations.

To train each one, the first N questions closest to the centroid of the cluster to which they belonged were taken. In the case of the classes assigned to a number of clusters $C >= 2$, the first $\lfloor N/C \rfloor$ questions of each one were taken, to obtain a balanced training dataset. This amount was defined individually according to the results obtained in each trained model, taking as the final N, the value that delivered the highest precision in the testing group.

With the intention that the classification model could categorize any type of question, training the model re-using the embedding created in the clustering process was discarded, since upon the arrival of a question outside the group used for training or testing , its characteristic representation could not be obtained. As a consequence, it was decided to encode the questions with two pre-trained neural embedding models: *bert_base_uncased*[12] and *all-MiniLM-L6-v2*[13]; both in the state of the art.

For the question typing process, the classifiers *Logistic Regression* and *Linear Suport Vector Classification (LinearSVC)* were tested; using an 80/20 division of the data set to train and test the resulting 4 combinations. The Figures (7), (8), (9) and (10) show confusion matrices, precision percentages and the N values used, for the combinations: **M1**: *'bert_base_uncased + Logistic Regression'*, **M2**: *'bert_base_uncased + LinearSVC '*, **M3**: *'all-MiniLM-L6-v2 + Logistic Regression'* and **M4**: *'all-MiniLM-L6-v2 + LinearSVC'* respectively.

Taking into account the previous results, although the four combinations achieved good precision, a slight increase in performance is observed in those combinations that used *LinearSVC* as a multi-class classification model. On the other hand, when comparing performances in relation to the embedding models used, the encodings with *bert_base_uncased* were superior. The latter, being attributable to the greater amount of information contained in the gen-

[12] https://huggingface.co/bert-base-uncased

[13] https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

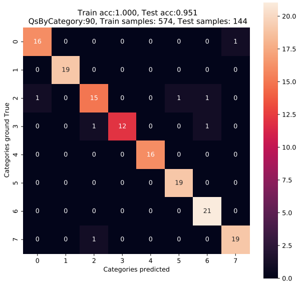

Fig. 7: Training results of M1: 'bert_base_uncased + Logistic Regression'.

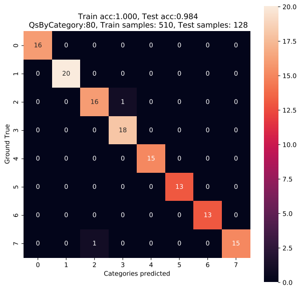

Fig. 8: Training results of M2: 'bert_base_uncased + LinearSVC'.

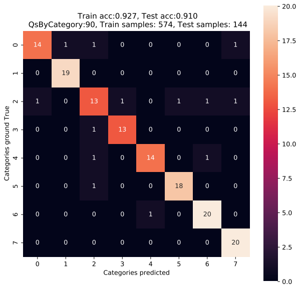

Fig. 9: Training results of M3: 'all-MiniLM-L6-v2 + Logistic Regression'.

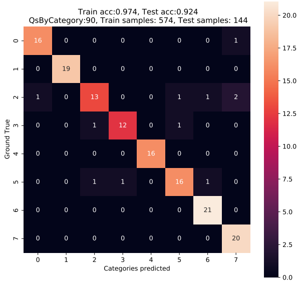

Fig. 10: Training results of M4: 'all-MiniLM-L6-v2 + LinearSVC'.

erated dimensionality vectors 768, against 384 for *all-MiniLM-L6-v2*. The Figure (11), shows the distribution of predictions on all the questions of the 17 analyzed clusters, using the M2 combination.

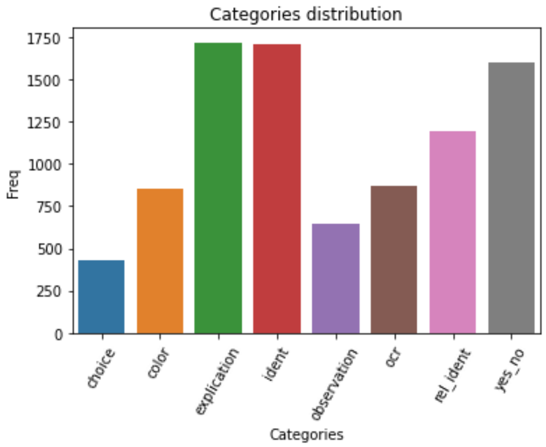

Fig. 11: Categories distribution using predictions of the models combination M2.

5 Models testing

In this last section, the results of the predictions obtained using the M2 combination (bert_base_uncased + LinearSVC) will be analyzed, which was the one that in general terms, with 98% precision in the testing group, delivered the best results. Unlike the previous section, Section (4), where prediction distributions were generated on the 9003 pre-processed questions, Figure (11); at this stage, the model was fed with all questions from the unfiltered VizWiz-VQA dataset. Again, the test set will be left aside, since it does not have associated answers and therefore categories to be able to contrast results.

Figure (12), shows the confusion matrix between the new classes: *choice*, *color*, *explanation*, *ident*, *observation*, *ocr*, *rel_ident*, *yes_no* , and the old categories: *yes/no, unanswerable, other, number*; over the total of 24842 questions. Note that each cell of the matrix represents what percentage of the old category ('Answer Type') was classified with the prediction indicated immediately below, on the horizontal axis.

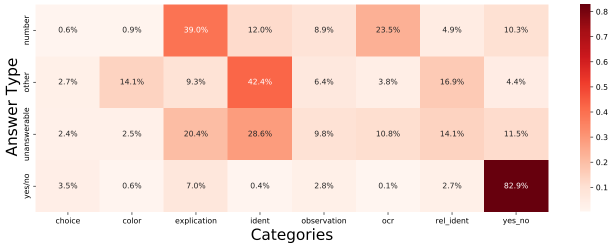

Fig. 12: Confusion matrix of M2 combination: 'bert_base_uncased + LinearSVC', over full VizWiz-VQA dataset.

Two natural observations that can initially be made on the matrix of Figure (12), are on the classes 'number' and 'yes/no'; unique descriptive original classes of the set. If we look at the 'yes/no' class, we can see that the ∼82% of the questions maintained their categorization, which gives a very good first impression, regarding the confidence of our classifier. When analyzing the class 'number', we see that the predictions are centered with the ∼23% and the ∼39% in the predictions 'ocr' and 'explication' respectively. Although it may seem wrong at first glance, this is quite consistent. Questions such as: *'How many cups are there?', 'How many doors?' or 'How many fingers in this image?'* require counting objects, and do not need optical character recognition capabilities such as: *'What is the number on this barcode?', 'What is the oven temperature setting?' or 'What is the second number?'.*

Being able to discover the conformation of the 'other' class was one of the main motivations that led to develop this work. You can see three large subclasses that characterize it: 'color', 'rel_ident' and 'ident' with the ∼14%, ∼16% and ∼42% respectively. Which indicates that this class mixes questions of identification of colors and identifications of direct and indirect objects, through characteristics or properties already known. Some examples of such groups are detailed in Table (3).

Answer Type	Category	Question
other	ident	What kind of soup is this?
other	ident	What are these?
other	ident	What is this product?
other	ident	What kind of pie is this?
other	rel_ident	What is in this jar?
other	rel_ident	what is in front of me?
other	rel_ident	What is against the wall next to the door?
other	rel_ident	What is on this package?
other	color	Is my shirt dirty and what color is it?
other	color	What color is this cover? Thank you
other	color	What color are my pants?
other	color	What is the color for this keyboard?
other	color	What color is this?

Table 3: Examples of predictions for old class 'other'.

Finally, when analyzing the class 'unanswerable', it is observed that the predictions with the highest percentages correspond to 'explication', 'yes_no' and again 'ident' and 'rel_ident'. Four categories that require a well-defined associated image and in context with the question asked, so that they can be correctly answered; something that does not characterize this dataset.

In Annex (C), 50 random predictions samples are detailed with each of eight new categories, made with the models of the M2 combination, on the VizWiz-VQA dataset.

6 Conclusions and Future directions

The main effort of this work was aimed at obtaining a disaggregation of the majority classes of the set of visual questions VizWiz-VQA. In a first stage, unsupervised techniques and different input data coding strategies were used, in order to feed the clustering algorithm with the embedding that would deliver the best results. In this part, the *KMeans* algorithm was used as a clustering method, and two embedding strategies: one based on the construction of occurrence matrices through n-grams + dimensionalid reduction, and the other, through pre-trained neural embedding models. such as *fastText* and *doc2Vec*.

After the clustering, a process of analysis of the results was started in search of properties and characteristics that would allow the natural identification of each one of the clusters obtained. As a consequence, a set of 8 new categories was proposed: *choice, color, explanation, ident, observation, ocr, rel_ident, yes_no*, laying the foundations for the development of the second phase of the project. The objective of identifying new classes to replace those already predefined in the original dataset was due to the fact that the latter are unspecific and do not allow in-depth knowledge of the type and nature of questions they contain.

With the new classes already defined, in the second stage, two classifier models were trained: *Logistic Regression* and *Linear Suport Vector Classification*. In addition, in order for the final model to be able to label any question outside of the training and testing groups used, the input data for such training (in this case, the questions) were coded by testing two pre-trained models of neural embedding, both in the state of the art: *bert_base_uncased* and *all-MiniLM-L6-v2*. As a result, after testing four combinations (embedding model, classification model), the tuple *bert_base_uncased + LinearSVC* was selected as the definitive model, giving ∼ 98% precision in the set of test questions.

Subsequently, with the classification model trained on the new proposed question categories, the expected disaggregation of the complete set of VizWiz-VQA could be carried out. The results obtained when labeling all the questions with the new 8 classes, allowed to know several important points.

The old classes 'yes/no' and 'number', were the two purest categorizations found. For the first one, almost 83% of the answers were binary. On the other hand, in the class 'number' it was observed that not only were there questions directly related to the recognition of numbers in some given environment, but also, questions related to object counts made up ∼ 39% of the group and not required OCR capabilities. With respect to the most emblematic classes, in 'other' the questions fell mainly into three subcategories, those related to the identification of objects being the most prominent, followed by relational questions and finally color identification. For the case of 'unanswerable', four main types of predictions were identified: 'explication', 'ident', 'rel_ident' and 'yes_no'. Of these, the first three are categories of questions that are very difficult to answer since they require very great reasoning skills, prior knowledge and management of perspective; if to this is added the characteristic poor qualities of the associated images, it would not be a surprise that these types of questions occupy this mysterious classification.

Finally, and thinking about possible future work, it is planned to use the results of the classification model to train a conditioned question and answer model (cQ&A). That is, with the re-classified questions, the new model will be fed, not only with the question of interest, but also with an extra conditioning, its category. In this way, and as the cGANs [5] (Conditional Generative Adversarial Networks) algorithms do, the result can be directed. This is very interesting since questions of the style *'Could you tell* *me what color is this?'*, which would trivially be answered with yes|no, when passing for example the category 'color', the model would be forced to hopefully it will return the name of a color, disambiguating the question to receive the specific type of response that is desired.

References

1. Ashok, A., Natarajan, G., Elmasri, R., Smith-Stvan, L.: SimsterQ: A similarity based clustering approach to opinion question answering. In: Proceedings of The 3rd Workshop on e-Commerce and NLP. pp. 69–76. Association for Computational Linguistics, Seattle, WA, USA (Jul 2020). https://doi.org/10.18653/v1/2020.ecnlp-1.11, https://aclanthology.org/2020.ecnlp-1.11
2. Bigham, J.P., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R.C., Miller, R., Tatarowicz, A., White, B., White, S., Yeh, T.: Vizwiz: nearly real-time answers to visual questions. In: Proceedings of the 23nd annual ACM symposium on User interface software and technology. pp. 333 – 342. UIST '10, ACM, ACM, New York, NY, USA (2010/// 2010). https://doi.org/10.1145/1866029.1866080, http://doi.acm.org/10.1145/1866029.1866080
3. Brady, E., Morris, M.R., Zhong, Y., White, S., Bigham, J.P.: Visual challenges in the everyday lives of blind people. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. p. 2117–2126. CHI '13, Association for Computing Machinery, New York, NY, USA (2013). https://doi.org/10.1145/2470654.2481291, https://doi.org/10.1145/2470654.2481291
4. Gurari, D., Li, Q., Stangl, A.J., Guo, A., Lin, C., Grauman, K., Luo, J., Bigham, J.P.: Vizwiz grand challenge: Answering visual questions from blind people. CoRR **abs/1802.08218** (2018), http://arxiv.org/abs/1802.08218
5. Mirza, M., Osindero, S.: Conditional generative adversarial nets. CoRR **abs/1411.1784** (2014), http://arxiv.org/abs/1411.1784
6. P, D.: MixKMeans: Clustering question-answer archives. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. pp. 1576–1585. Association for Computational Linguistics, Austin, Texas (Nov 2016). https://doi.org/10.18653/v1/D16-1164, https://aclanthology.org/D16-1164
7. Terao, K., Tamaki, T., Raytchev, B., Kaneda, K., Satoh, S.: An entropy clustering approach for assessing visual question difficulty. IEEE Access **8**, 180633–180645 (2020). https://doi.org/10.1109/ACCESS.2020.3022063

A List of conversational sequences

'can you tell me', 'can you tell', 'tell me', 'please tell me', 'please read', 'can you please tell me', 'please can you tell me', 'could you please tell me', 'can you please', 'can you read to me', 'can you please read', 'can you see','can you read','can you give me','can you help me','are you able to', 'i do not know if this', 'i want to know the'.

B Top 20 questions from each cluster, from best strategy selected (Qst+BestAns).

Question	Best Answer	Distance
what color are those pants?	pink	0.213182
what color is cup?	yellow	0.230284
>what color is this product?	pink	0.237367
what color is this item?	grey	0.242185
what color is this device?	silver	0.248018
what color is this phone case?	brown	0.248018
what color is this image?	blue	0.248018
what color is this woman's top?	blue	0.254964
what color is this object?	green	0.254964
what color is this button?	clear yellow	0.254964
what color is this table?	brown	0.254964
what color is this cup?	blue	0.254964
what color is this dog?	brown	0.254964
what color is this gift bag?	pink	0.254964
what color is this man's pants.	grey	0.254964
what color is my book bag?	blue grey	0.262036
what color is this top ?	pink	0.263126
what color is this glass?	blue	0.263126
what color is this case?	green	0.263126
what color is this pant?	red	0.263126

Table 4: [Cluster 0] — 20 top questions

Question	Best Answer	Distance
what does this cup say.	1 3 cup	0.241263
what does this label say?	green pigeon	0.248653
what does this box look like?	phone	0.257584
what does this plant look like?	plant	0.261638
what does this dress look like?	curtain	0.266728
what does this item look like?	necklace	0.266728
what does this say	intel pentium	0.267346
what does this packet say?	hot cocoa mix	0.267346
what does this say?	8	0.267346
what does this box say?	toad training	0.267346
what does this cup look like?	movie poster	0.272992
what does the label say?	shiner rye	0.27753
what does this food label say?	uncle bens	0.278919
what does this character say?	3	0.278919
what does this label say.	new bothwell mb	0.278919
what does this box say on top?	mary kay	0.280023
what does this pregnancy test show?	1 line	0.280586
what does this monitor look like?	blank	0.280586
what does he look like?	dog	0.281864
what does the sky look like?	dark	0.282237

Table 5: [Cluster 1] — 20 top questions

Question	Best Answer	Distance
what is this item?	solar garden light	0.137387
what is this wine?	wine	0.137387
what is this canned good?	green beans	0.137584
what is this device?	cell phone	0.137584
what is this bag look like	suitcase	0.137584
what is this drink?	coca cola	0.13922
please tell me what is this box?	2 spicy bean burgers	0.13922
what is this thing right here?	pepperoni pizza	0.140246
what is this box	mcdonalds	0.140246
what is this product key?	021 08454	0.140246
what is this soda called?	pepsi	0.140246
what is this bill?	10	0.140246
what is this label?	hunts pasta sauce	0.140246
what is this box?	coconut sponge	0.140246
what is this box from?	powerskin	0.140246
what is this person wearing?	blue shorts	0.140246
what is this ready meal package?	chicken	0.140246
what is this spice?	apple pie spice	0.142654
what is this cup?	coffee	0.145769
what is this pink packet	sweet n low	0.145769

Table 6: [Cluster 2] — 20 top questions

Question	Best Answer	Distance
what is the sky look like today?	clear	0.190297
what is the oven temperature set at?	230	0.191961
what is the brand name?	clover organic farms	0.194766
what is the temperature set at?	75	0.198866
what is the oven temperature control setup?	dial	0.204423
what is the oven control temperature?	350	0.211605
what is the oven temperatures setting?	325	0.211605
what is the name of ths product	oxibooster	0.212719
what is the sodium content?	10 mg	0.220584
what is the product?	cheese sticks	0.220584
what is the dial set at?	450	0.220584
what is the name of the mouse brand?	microsoft	0.22708
what is the name of the cd	patsy cline	0.22708
what is the name of the water?	naturliches	0.227491
what is the name of the drink?	irn bru	0.227491
what is the name of the story?	sleep book	0.227491
what is the name of the flower?	carnation	0.227491
what is the name of the magazine?	popular mechanics	0.228351
what is the name of the restaurant?	aladdin natural eatery	0.229712

Table 7: [Cluster 3] — 20 top questions

Question	Best Answer	Distance
can you read who this christmas card is from?	unanswerable	0.216867
where is this box from?	unanswerable	0.221511
what food is inside this packet?	unanswerable	0.225636
where is this nut from?	unanswerable	0.230656
where is this coin from	unanswerable	0.230656
what temperature is this dial set too?	unanswerable	0.231024
who is this parcel from?	unanswerable	0.237108
where is this from?	unanswerable	0.237108
when is this card?	unanswerable	0.237108
who is this from?	unanswerable	0.237108
what size is this item?	unanswerable	0.237371
where is this table?	unanswerable	0.245022
who is this dog?	unanswerable	0.245022
why is this computer not booting up?	unanswerable	0.254572
who is this person?	unanswerable	0.254572
what temperature is this thermostat set to?	unanswerable	0.260985
what size of cereal is this box.	unanswerable	0.262043
what temperature is this oven set to	unanswerable	0.263272
who is this character?	unanswerable	0.265953
where is this made?	unanswerable	0.265953

Table 8: [Cluster 4] — 20 top questions

Question	Best Answer	Distance
is this candy or chocolate?	chocolate	0.196788
is this regular or decaf coffee?	regular	0.196788
is this fire engine red or yellow	red	0.196788
is this blue or purple?	blue	0.19726
is this regular or caffeine free?	caffeine free	0.19726
is this regular mountain dew or diet?	regular	0.19726
is this shower gel or lotion?	gel	0.19726
is this catalina regular or catalina free?	free	0.19726
is this sweatshirt brown or tan?	brown	0.19726
is this diet or regular pepsi?	diet	0.19726
is this diet pepsi, regular, or caffeine free?	diet	0.19726
is this cloth pink or blue?	pink	0.19726
is this flowers or stripes?	stripes	0.19726
is this yarn blue or purple?	purple	0.197804
is this stripes or flowers?	flowers	0.197804
is this shampoo, conditioner, or lotion?	conditioner	0.197804
is this decaf or regular coffee?	regular	0.197804
is this shampoo or conditioner?	shampoo	0.197804
is this with chocolate or with fruit?	chocolate	0.197804
is this inhaler blue or yellow?	yellow	0.197804

Table 9: [Cluster 5] — 20 top questions

Question	Best Answer	Distance
what dollar bill is this?	5	0.165812
what denomination is this?	20	0.165812
what temperature is this?	450	0.16603
what brand is this>	trojan	0.16603
what brand is this?	winston	0.16603
what schwan's dinner is this?	beef shepherds pie	0.167918
what harry potter book is this?	sorcerers stone	0.167918
what bill is this?	1 dollar	0.168106
what product is this?	cereal almonds	0.17162
what bill denomination is this?	1 dollar	0.17162
what dollar amount is this?	5	0.172827
what dinner is this?	beef strips	0.172827
what card is this?	justice	0.172827
what video game is this?	tiger woods pga tour 10	0.172827
what gift card is this?	dunkin donuts	0.172827
what gift card is this	tim hortons	0.172827
what cleaning product is this?	409 all purpose cleaner	0.172827
what food is this?	pizza	0.172827
what product is this	coffee	0.17621
what fruit is this?	apple	0.181007

Table 10: [Cluster 6] — 20 top questions

Question	Best Answer	Distance
what product is this, including brand name if possible?	unsuitable	0.239348
what brand is this bag?	unsuitable	0.243779
what brand is this radio?	unsuitable	0.262318
what exactly is this product?	unsuitable	0.262318
what temperature this thermometer is set on?	unsuitable	0.269034
what brand is this shaver?	unsuitable	0.279981
can you tell me what is inside this box?	unsuitable	0.281408
what discount is written on this card?	unsuitable	0.284921
what temperature is this thermometer on?	unsuitable	0.288678
what denomination is this dollar bill?	unsuitable	0.290737
can you tell me what the oven might be set on?	unsuitable	0.297323
what items are check marked on this card?	unsuitable	0.301576
what product is that?	unsuitable	0.302399
what model is this keyboard?	unsuitable	0.302954
who is this letter from?	unsuitable	0.304477
what size is this shirt?	unsuitable	0.305887
what temperature is my oven at?	unsuitable	0.310959
what book is this, thank you.	unsuitable	0.311527
can you read what is written on this box?	unsuitable	0.312826
if this is any better.	unsuitable	0.314554

Table 11: [Cluster 7] — 20 top questions

Question	Best Answer	Distance
what kind of coffee is this?	house blend	0.189632
what kind of soup is this?	chicken tortilla	0.189748
what kind of soda is this	pepsi	0.189748
what kind of k-cup is this?	english breakfast tea	0.190818
what kind of drink is this?	dr pepper	0.190818
what kind of dog is this	golden retriever	0.192979
what kind of food product is this?	corn chips	0.192979
what kind of dog is this?	golden retriever	0.192979
what kind of cat food is this?	meow mix	0.192979
what kind of dog food is this?	australian lamb	0.192979
what kind of tv dinner is this?>	escalloped chicken noodles	0.192979
what kind of soft drink is this?	dr pepper	0.196369
what kind of drink is this	diet sunkist	0.20113
what kind of frozen dinner is this?	baked chicken	0.20113
what kind of ice cream is this?	vanilla bean	0.20113
what kind of food is this?	rice	0.20113
what kind of tassimo coffee is this?	house blend	0.20113
what kind of dinner is this?	3 cheese tortellini	0.20113
what kind of keurig cup is this?	hot apple cider	0.207401
what kind of tv dinner is this?	lean cuisine	0.207401

Table 12: [Cluster 8] — 20 top questions

Question	Best Answer	Distance
hello hello hello.	unanswerable	0.154899
test one two, test one two	unanswerable	0.160078
expiration date?	unanswerable	0.163961
test test.	unanswerable	0.163961
business card.	unanswerable	0.17366
help with question.	unanswerable	0.17366
new app test.	unanswerable	0.17366
just testing.	unanswerable	0.17366
just answer anything.	unanswerable	0.187578
record record	unanswerable	0.187578
test question.	unanswerable	0.187578
can you tell now?	unanswerable	0.187578
testing your phone	unanswerable	0.206207
cooking directions?	unanswerable	0.206207
read directions.	unanswerable	0.206207
cooking directions.	unanswerable	0.206207
testing one two three testing.	unanswerable	0.206207
so much.	unanswerable	0.206207
hello computer	unanswerable	0.206207
testing, testing	unanswerable	0.206207

Table 13: [Cluster 9] — 20 top questions

Question	Best Answer	Distance
describe this item.	bath shower gel	0.186666
identify this product.	brown sugar	0.186666
can you please describe this card?	green person holding ...	0.186666
can you please describe this label?	lotion	0.19587
read this label.	dermacol acne clear	0.203199
identify this object	purse	0.212775
who wrote this book?	john green	0.212775
can you please identify this tin?	bug spray	0.212775
identify this object.	granola bar	0.212775
can you tell who put this one out?	top chef	0.212775
this box.	tea light candles	0.224977
describe this dress.	short	0.224977
can you tell if this has one line or two lines?	1 line	0.245812
if this green beans or kidney beans?	green beans	0.248096
que es	coca cola	0.249528
sky look like.	cloudy	0.249528
please describe this gift card	bath body works	0.250951
this candle	yellow	0.259137
can you give me information about this bar code?	1284353636	0.259137
name this object.	shaving cream	0.259137

Table 14: [Cluster 10] — 20 top questions

Question	Best Answer	Distance
what brand is this pop?	coca cola	0.217042
what temperature is this set at?	500	0.221288
what denomination is this note?	1 dollar	0.224423
what button is mountain dew?	5	0.227628
what temperature is this cooked at?	400	0.228653
what brand is this coffee?	whittard	0.234139
what product is this made by?	amys	0.234139
what denomination is this bill?	20	0.234139
what brand is this lotion?	secret charm	0.234139
what brand is this mouse?	dell	0.241058
what brand is this camera?	canon	0.241058
what brand is this hand sanitizer?	purell	0.241058
what denomination is this money?	20	0.241058
what countries are we looking at?	canada usa mexico	0.24214
what brand is this recorder?	olympus	0.249607
what brand is this popcorn?	act ii	0.249607
what video game is this one?	mortal kombat	0.249607
what time is this play?	18:37	0.272508
what scent is this lotion?	sweet pea	0.272508
what flavor is this pasta sauce?	smoked bacon tomato	0.280009

Table 15: [Cluster 11] — 20 top questions

Question	Best Answer	Distance
can you tell me if this tube is suntan lotion?	yes	0.2342
can you tell me if this cat is cute?	yes	0.25548
is this box right side up?	yes	0.262611
can you tell me if this looks like hamburger?	yes	0.273877
is this right side up?	yes	0.275559
are those piano keys?	yes	0.281417
can you tell if this is broccoli cheese soup?	yes	0.29859
is this an apple product?	yes	0.300372
is this an orange sim card?	yes	0.300372
can you see this image?	yes	0.303456
is this an iphone?	yes	0.312127
is this an orange?	yes	0.312127
is this remote control?	yes	0.312127
is this shampoo?	yes	0.312127
is this the new apple keyboard?	yes	0.316064
is he fat?	yes	0.322635
can you see if there are roots growing?	yes	0.323669
is there any writing on this dressing packet?	yes	0.324173
is this a 20 dollar bill?	yes	0.324667
is this a violent video game?	yes	0.324667

Table 16: [Cluster 12] — 20 top questions

Question	Best Answer	Distance
can you tell me who this card is from?	no	0.262782
are you able to read this business card?	no	0.273327
is there any writing on here?	no	0.294614
this piece of paper?	no	0.300474
is someone standing at the door?	no	0.3015
can you read this label	no	0.310389
can you read this paper?	no	0.310389
can you read this label?	no	0.310389
can you tell when this pack expires?	no	0.310389
any copying instructions.	no	0.314089
can you see any text?	no	0.314089
is there an expiration date?	no	0.316663
is there an expiration date	no	0.316663
can you tell if the soup?	no	0.317849
hi, can you see the cooking instructions for this dish?	no	0.319084
is there anything on this page?	no	0.325242
do the clouds look like storm clouds?	no	0.327964
are the directions visible now?	no	0.32901
are the directions showing now?	no	0.32901
any water bugs or anything?	no	0.330578

Table 17: [Cluster 13] — 20 top questions

Question	Best Answer	Distance
what is inside this canned good?	corn	0.209963
what is inside this box?	salisbury steaks	0.215137
what is inside this image?	leg	0.221753
what is an iphone?	cell phone	0.23364
what is written in here?	juice pack air	0.247832
what is inside?	tea	0.248061
what is written on this label?	fish oil	0.251888
what is written on this box?	dunhill	0.254562
can you tell me what is written on this card?	7259 7694	0.254562
what is that product?	cleaning product	0.261805
what am i looking at right now?	beer	0.264117
please tell me what is in this box	childrens medical box	0.268764
what is written on this tube?	usher after shave	0.269897
what is in this box?	beef stroganoff	0.274292
what is in this box	roast chicken	0.274292
can you see what is in this package?	chicken	0.274292
what is inside this can?	soup	0.275462
what is in this package	pumpkin pie spices	0.276746
can you tell me what is in this box?	mixed nuts	0.276746
tell me what is in this box.	spaghetti meatballs	0.276746

Table 18: [Cluster 14] — 20 top questions

Question	Best Answer	Distance
can you tell me about the bag?	blue	0.27387
can you please describe the towel?	grey	0.27387
where is the red car?	top right	0.287574
where is the menu button?	bottom left	0.292444
when is the expiration date?	feb 21 2014	0.294971
where is the sky?	up	0.303601
where is the dog?	on floor	0.313836
is the painting right side up or upside down?	right side up	0.314202
the sky look like?	cloudy	0.317186
you read the highlighted text.	top played games	0.320791
where is the coffee?	desk	0.325309
which one is the blue one?	right	0.326312
describe the photo.	guy in chair	0.330203
i am trying to get the expiration date on this milk.	feb 6 12	0.330301
what temperature is showing on the display?	0	0.331315
which is the diet coke button?	top button	0.331805
where is the printer?	on table	0.333687
where are the keys?	on towel	0.333687
is the light on or off?	off	0.334325
which one is the diet coke?	far right	0.338053

Table 19: [Cluster 15] — 20 top questions

Question	Best Answer	Distance
what is the expiration date?	unanswerable	0.180796
what are the cooking instructions - microwave?	unanswerable	0.181282
what is the cooking instructions?	unanswerable	0.184517
what is the top chef question?	unanswerable	0.229252
what are the instructions for using this product?	unanswerable	0.229273
what are the cooking directions for this box?	unanswerable	0.22968
what are the directions for this product?	unanswerable	0.22968
what are the cooking instructions for this packet?	unanswerable	0.233524
what are the cooking instructions for this item?	unanswerable	0.233524
what are the administration instructions for this product?	unanswerable	0.235919
what is the expiration date of this almond milk?	unanswerable	0.238334
what is the expiration date of this milk?	unanswerable	0.238334
what is the name of this cd?	unanswerable	0.238334
what is the name of this item?	unanswerable	0.238776
what is the name of this menu?	unanswerable	0.238776
what is the expiration date of this yogurt?	unanswerable	0.238776
what is the name of this product ?	unanswerable	0.239539
what is the expiration date of this turkey?	unanswerable	0.239584
what is the name of this talking book?	unanswerable	0.240796
what is the brand name of this air conditioner?	unanswerable	0.240796

Table 20: [Cluster 16] — 20 top questions

C 50 random predictions using combination M2, over full VizWiz-VQA dataset.

Answer Type	**Category**	**Question**
other	rel_ident	If i zoom in can you try to read them or is it just too small?
yes/no	yes_no	Are these strawberries?
unanswerable	color	What color is this blanket?
unanswerable	ident	What is this item?
unanswerable	ocr	What is the name of this drink?
other	choice	Okay this is my last try with this. Could you please tell me what the color of this outfit is, that she's wearing?
other	color	What color is this bell?
other	explication	Which headphone is the pink one. The one on the left or the one on the right.
other	color	WHat color is this?
other	ident	Okay I need to know what this is and I definitely know it's not chicken fillets.
other	ident	What is this medication?
unanswerable	explication	Yes I find this in apartment and I am from foreign country and I don't kno what it is.
other	rel_ident	What is this picture? What is this picture?
other	explication	Can you tell me the serving size and calories, please.
yes/no	yes_no	Can you see if there are roots growing?
unanswerable	ident	What flavor is this?
unanswerable	rel_ident	What's in this box?
other	ident	what's is this item?
yes/no	rel_ident	Is there any other writing other than Stove Top? I need to know what is in this package. Thank you.
other	color	What color is this?
unanswerable	ocr	we called what is the flavor
unanswerable	explication	What page number is this above? Thank you.
unanswerable	ident	What is this?
other	color	What color is this?
yes/no	choice	Is my light on? And I have a question, how late do you have to work tonight, I'm just curious
other	ident	What is this?
other	ocr	Alright what is the expiration on this carton of milk?
other	ident	What is this can?
unanswerable	ident	What is this?
other	ident	Whats this?
other	observation	What does the display say?
other	observation	What does it say?
other	ident	What wine is this?
other	ident	What's this?
other	yes_no	Can you tell what this is?
other	color	What color is this?
other	yes_no	What is this a picture of? Can you tell me?
yes/no	yes_no	Is there caffeine in there?
other	rel_ident	What is that?
unanswerable	ocr	What is the name of this product?
yes/no	yes_no	This is an advertisement?
other	ident	What is this?
other	ident	What is this spice?
unanswerable	yes_no	Can you tell me how to make this in the microwave?
other	rel_ident	What is in this picture?
other	rel_ident	What is in this can please?
unanswerable	observation	What does the label on this say?
unanswerable	observation	For how long do I cook this in the microwave?
other	rel_ident	What's on this channel?
other	explication	can you tell the name of this product if possible, please? thanks

Table 21