**Political Advertisements and Gun Usage in the 2020 Election Cycle**

Brigitte Goeler-Slough, Chiara Christie, Emma Tuhabonye

QAC239 - Final Project

December 16, 2022

# Introduction

For decades, gun images have been used by both parties in campaign advertisements. However, in recent years, there has been an increase in mass shootings and gun violence that has encouraged a bipartisan discourse surrounding gun laws. This has prompted cultural and political use of guns to drive partisan arguments. As a result, researchers have focused on analyzing political ads as a way to examine the discourse in reference to these issues.This is important in understanding the implications of the frequency of gun imagery in political ads, and the topics of these ads. Is there a mass influx of political ads discussing gun control and showing gun imagery? If so, this could open up more areas of research discussing the reasoning for this uptick. Having a way to track gun imagery and gun-related political ad trends may offer critical insights on how candidates view the importance of gun policy in each election cycle. This data potentially informs future studies looking into gun-related ad campaign trends.

A study from 2018 by Dean Colleen L. Barry and her team investigated the mentions of guns in candidate-related television and airings over four election cycles. Their data consisted of political advertisements from 2012-2018 election cycles (2012, 2014, 2016, 2018) and consisted of over 14 million ads. The team found that the share of these ads referring to guns increased 7% from 2012-18. “Pro-gun rights content dropped from 86% in 2012 to 45% in the 2018 cycle. Ads in favor of gun regulation and against the NRA increased over this time period” (Barry et al., 2020) The insights gained from this study inform our topic, as our data is from the 2020 election cycle, and would illustrate a trend that may reflect the Barry study.

Our dataset was TV advertisements sponsored by federal candidates for the 2020 general election. The project uses it to focus on the detection of gun images and the mention of guns in political television ads. We used deep learning techniques such as image classification to analyze videos in these political ads. Taking this data and merging it with transcriptions of the ads themselves, we then used sentiment analysis models on the speech data to identify how the candidates were discussing gun control (pro vs. anti) based on the connotations of the words that they chose to use in the ad. We are looking to investigate how candidates in the 2020 election cycle discuss gun laws. Our aim is to create a model that can parse through ads to identify guns, train a sentiment classifier on gun control political ads and create a model that can be used in future election cycles to identify subject matter trends. We hope that with this combination we can use ad transcriptions as well as gun imagery to identify candidates’ stances on gun control.

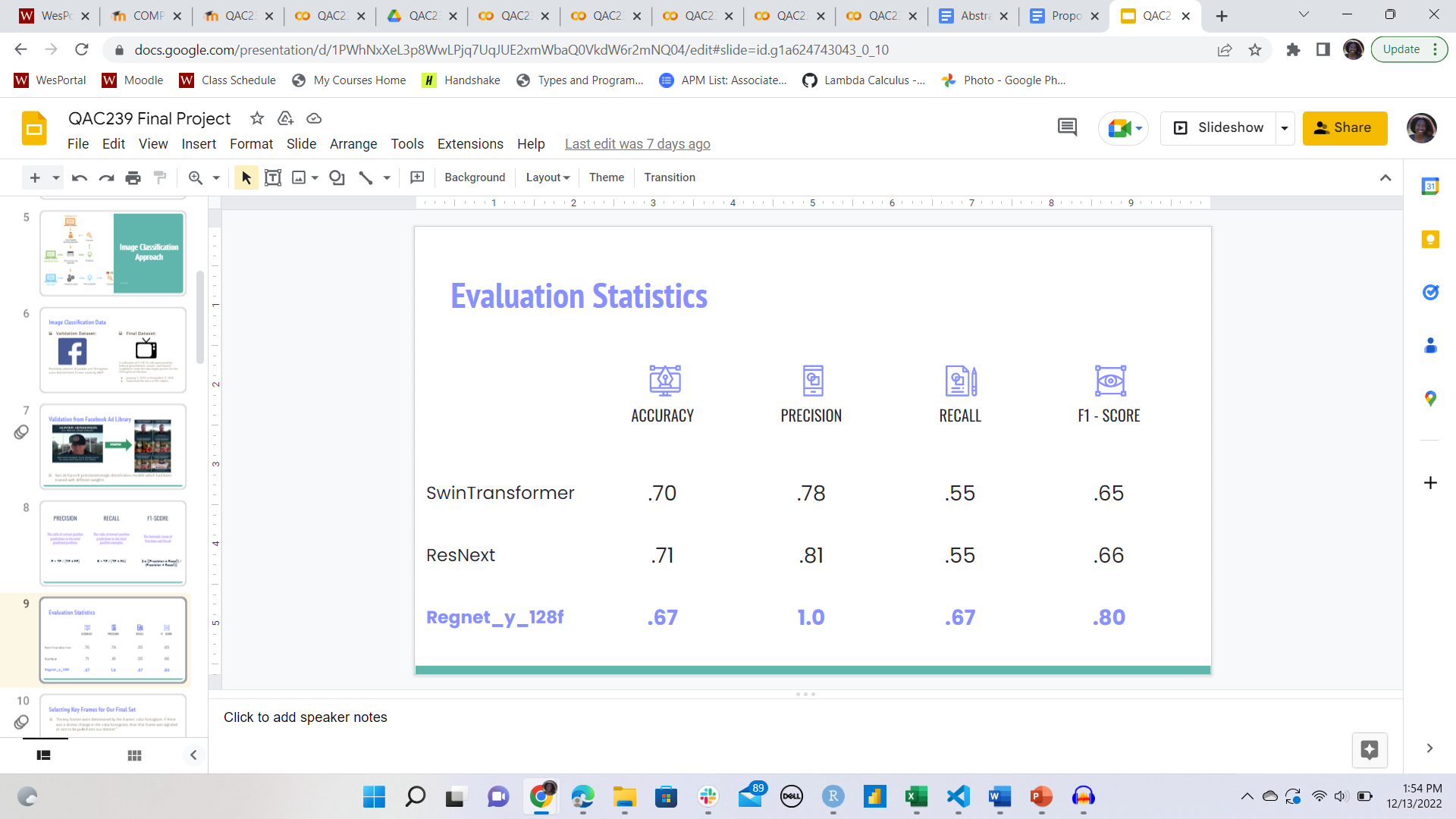
# Data & Methods

## Image Classification

Two dataset were collected – a validation set and our final set that the model was applied on. The validation set consisted of videos from Facebook Ad library and human coded results provided by the Wesleyan Media Project. We randomly selected 38 positive cases and 38 negative cases. We used FFmpeg to extract every single frame from each ad. After obtaining the frames, they were run through 20 Pytorch pretrained image classification models. Three top models were selected and compared. Our top 3 models included SwinTransformer, ResNext, and Regnet\_y\_128f.

The SwinTransformer is a hierarchical transformer whose representation is a computer with shifted windows. The shifted windowing scheme brings greater efficiency by limiting self-attention computation to non-overlapping local windows while also allowing for cross-window connection (Liu et al., 2021). ResNext model is a modularized network architecture for image classification. The network is constructed by repeating a building block that aggregates a set of transformations with the same topology (Xie et al. 2017). The RegNet\_y\_128f has an architecture from a network design space that parametrizes populations of networks. The process of the network design is similar to the classic manual design of networks, but more elevated (Xie et al. 2017).

Based on our evaluation statistics, we were able to conclude that RegNet\_y\_128f was our best model (*Table 1*). From there we decided to apply it to our final television set.



*Table 1*: Evaluation Statistics of the top 3 models.

Although the model had the lowest accuracy, the F1-score was the highest of the 3. The amount of correctly identified guns out of to the amount of predicted guns is 100%, the amount of guns that were correctly identified out of to the amount of actual guns in the ads was 67%. And F1 is the harmonic mean of those and came out to be 80%

The final dataset consists of a collection of 2,130 television ads sponsored by federal candidates from Democrats and Republicans for the 2020 general election. This dataset was also human coded by the Wesleyan Media Project where we got the tone of the ad and whether or not gun issues were mentioned. The television ads were collected from January 1, 2019 to November 3, 2020. After acquiring the television ads, we subsetted the data further to only include videos where gun issues were mentioned. This resulted in a dataset of 552 television advertisements.

The length of time it took to extract every frame and run that through the model was very large so we decided to experiment with selecting key frames from our final dataset. The key frames were determined by the frames’ color histogram. If there was a drastic change in the color histogram, then that frame was signaled as one to be pulled out of the dataset. To validate using the video summarization technique, we took out 85 videos and human coded them. There were 29 positive cases and 56 negative cases. On our human coded data, we calculated the precision, recall, and f1 score. With these results, the RegNet\_y\_128f model was applied on the final set and sent for sentiment analysis.

## Sentiment Analysis

The dataset we used for the sentiment analysis was the same dataset of 2020 TV ads, which had two human coded columns which were important for our analysis, which identified if the candidate was pro or anti gun control. The two columns depicting gun control position were merged into one column to determine the candidate’s view: (0 = anti-gun control, 1 = pro-gun control). We additionally had a dataset that included transcript data from the ads that used Google’s Speech-to-text to transcribe them. We merged these datasets, only selecting ads with a transcript confidence greater than 0.8, and selected the ads with the human coded pro or anti gun control, which resulted in 500 entries total (142 pro-gun control, 358 anti-gun control). We also wanted to look specifically at ads that our image classification model determined contained an image of a gun, to see how this would affect our model’s prediction of the sentiment of the ad, and this left us with 142 total entries (35 pro-gun control, 107 anti-gun control).

Our first attempt at sentiment analysis was to try a simple logistic regression classification on the data. We first had to vectorize the data to turn the transcript into a document-term matrix, and here we added a parameter to remove stopwords from the data (as they would not give us much information about the sentiment of the ad). We then performed a 80/20 train/test split of the data and ran our logistic regression classification on the data.

Given that our dataset was relatively small, it’s possible that our first random split was not a good representative split of the data. So to try to get a better understanding of how our model performs on the entire dataset, we decided to run 10-fold cross validation. This involves splitting the dataset into 10 consecutive folds, training and testing the model on each of these different folds to determine the average accuracy of the model on the different splits.

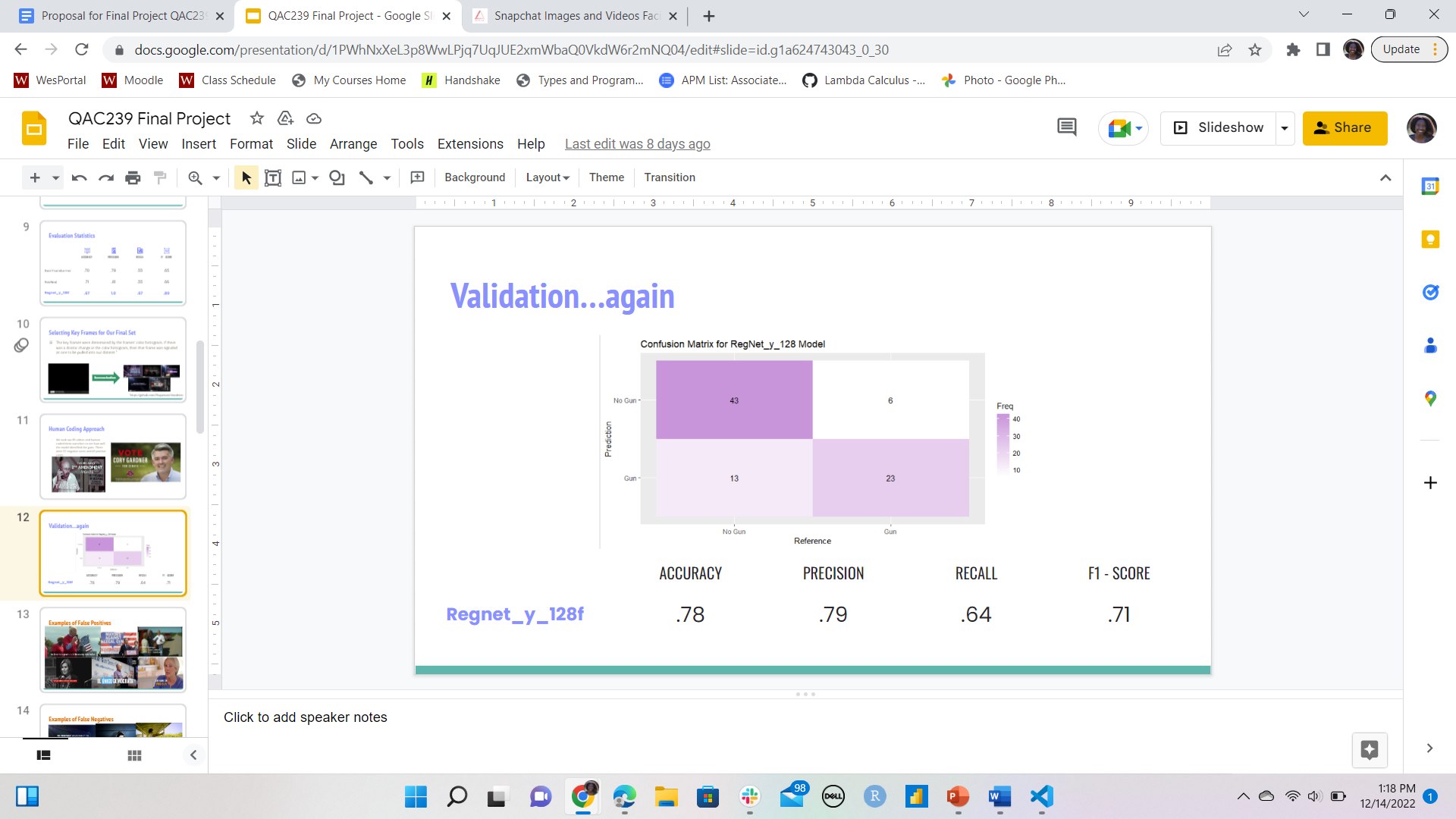
We also tried to optimize the performance of our model by trying out different types of models on the data to see if this could change the accuracy of our model. To optimize this even further, we also varied the hyperparameters to get the optimal results. We first tried Gaussian Naive Bayes; it assumes that all features are independent, and calculates the most likely class by calculating the probability of each class using the equation: and then chooses the class that has the highest probability. However, since this model has no parameters that can be changed, we weren’t able to tune it to optimize its accuracy. Next we tried a Random Forest model, which is essentially a decision tree which partitions the data best to predict the outcome, and adds bootstrap aggregating, which involves taking other samples from the dataset and averages the model results on these different samples, which reduces variance in the model (but if different subsets are sampled multiple times it can increase bias). In this model, the hyperparameter we can tune is the number of trees, so we ran the model varying the number of trees in intervals of 100 from 100 to 500 trees. Finally, we also tried a Linear Support Vector Machine, which tries to transform the feature space to fit the boundaries between the different possible classes. Here we could tune both C and gamma, so we did a grid search testing C at to 10, 100, and 1000 and gamma at 0.001, 0.005, 0.01, 0.05, and 0.1.

We were also curious to see how a pre-trained model would perform compared to our other classification models that work with the labeled training data that we feed it. For a pre-trained model, we decided to use Valence Aware Dictionary and sEntiment Reasoner (VADER). It was designed to primarily work in social media settings, but can work in general in many contexts. Unlike our previous models where we need to feed labeled training data, VADER is a lexicon and rule-based sentiment analysis model, where it can be run on unlabeled data. It has a lexicon of specific words with a corresponding sentiment score (positive or negative value), so after adding up all the values associated with the words in a given block of text, the model can determine the sentiment. This lexicon has been built using Amazon’s Mechanical Turk, which is a crowdsourcing platform, meaning that it is essentially a human coded sentiment score for different words, with many different people contributing to the value. VADER returns a positive, neutral, and negative value (the proportion of words that fall in each category) and a compound score which tells you the overall sentiment of the text, normalized to a value between -1 and 1. So if we consider a compound value greater than 0 positive/pro tone and less than 0 negative/anti tone, we can use VADER to predict if the overall sentiment of a transcript is pro or anti-gun control. We ran VADER on our entire dataset to see how it did on all of our data, and then also on the subset just including the ads which we predicted had an image of a gun in it.

After this, we wanted to compare how the ad actually containing the image of a gun may affect how our model predicted sentiment of the ad, so we chose to subset the data even further, now only including the ads that our image classification model predicted contained a gun. Since we found that our logistic regression model was one of the most successful models for the data, we decided to use this model with our new subset of the data to see how it would compare to the larger dataset, again using a 10-fold cross validation, as this dataset was even smaller than before so the chances of the random split not being representative were even greater.

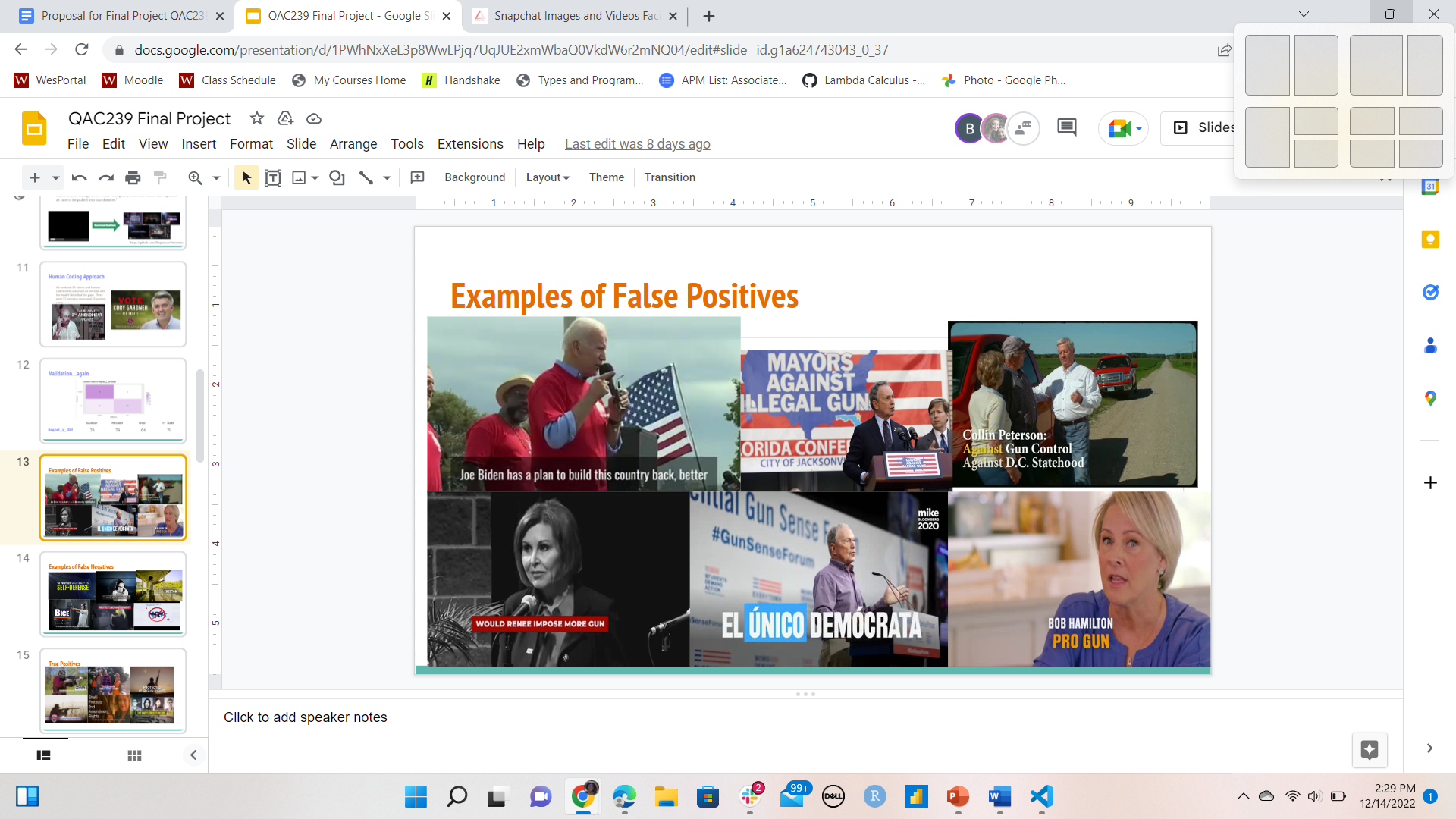
# Results

## Image Classifications



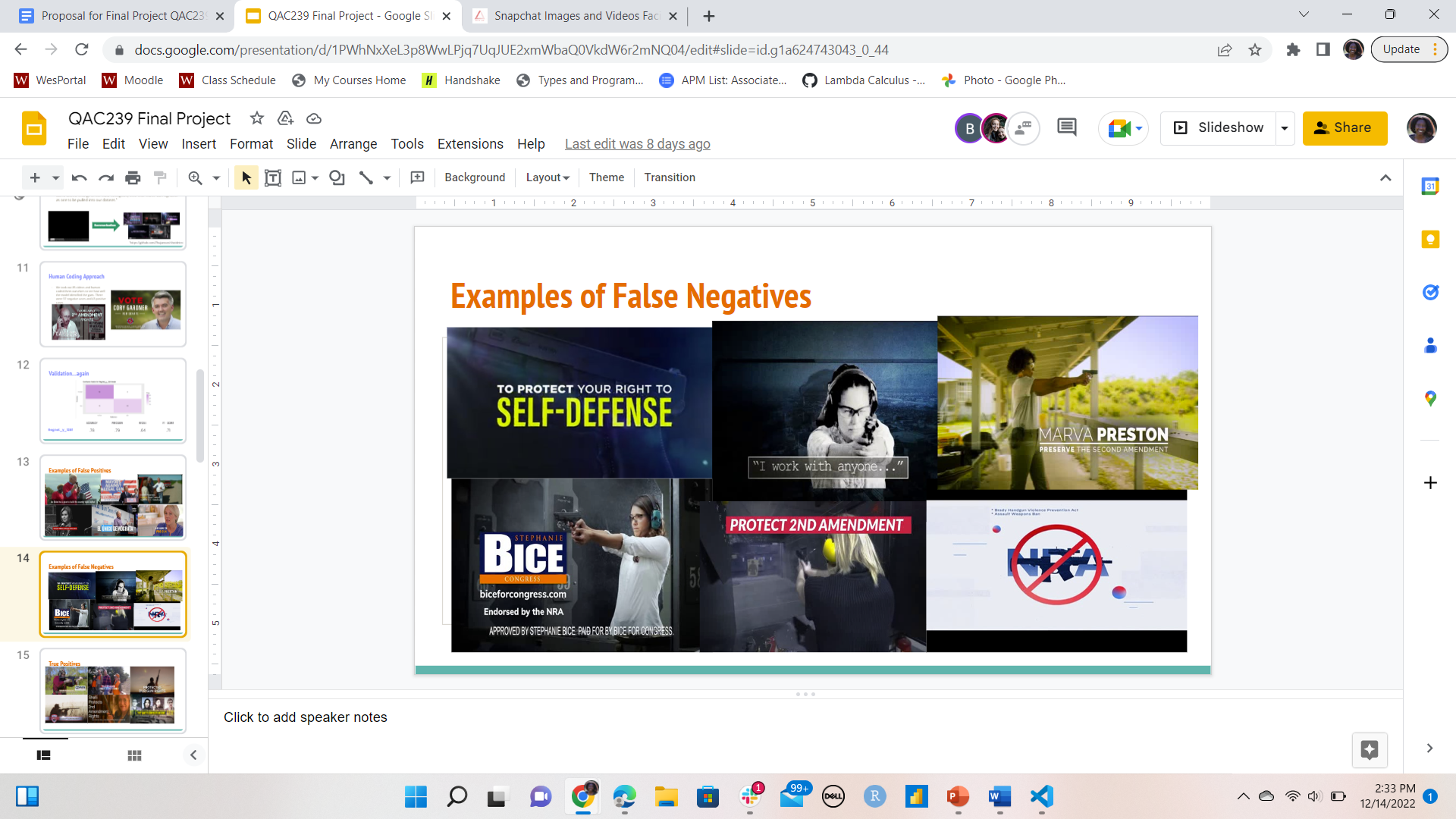
*Figure 1:* Confusion matrix and Evaluation Statistics on Human Coded Key Frames

The algorithm has an accuracy of 78% for classifying images with guns. The precision value of 0.79 for gun positive images shows that when our classification model predicts an ad has a gun, it is correct 79% of the time. The recall value of 0.64 reflects the percentage of gun images that were correctly identified by the RegNet\_y\_128f classifier out of all the images with guns.



*Figure 2:* Examples of False Positives

When we examined the mistakes that the classifier made, we saw that most of the images had a microphone in them. We were unsure if the classifier was mistaking a microphone for a gun. However, every image in the example in Figure 2 had the word “Gun” so we believe that the classifier returned images that had the word gun even if there was not an actual gun in the image.



*Figure 3:* Examples of False Negatives

Most of the false negative images seem to be because of the low contrast between the gun and the background or the person holding the gun. We also saw examples of the guns surrounded by a lot of colors and words and we suspect that made it harder for the gun to be identified.

## Sentiment Analysis

|  |  |
| --- | --- |

*Figure 4: Confusion matrix and classification report for Logistic Regression.*

From the confusion matrix and classification report for the logistic regression (for the 0.2 test split), it shows that with gun control issue ads, the model is predicting anti-gun control the majority of the time. A precision value of 0.91 shows that when our model predicts anti-gun control, it is correct 91% of the time, which is very high. When our model predicts pro-gun control, it is correct 77% of the time. The recall score of 0.93 means that the model correctly identifies 93% of all anti-gun control ads, compared to 71% for pro-gun control. The lower recall score for pro gun control ads may be due to the larger proportion of anti-gun control cases in the dataset, and therefore the training set. Models trained on unbalanced datasets have a hard time generalizing and predicting unseen observations and therefore may predict more anti-gun control cases as pro-gun control.

|  | Predicted = 0 | Predicted = 1 |
| --- | --- | --- |
| Actual = 0 | America greatest country Earth hard work. They respect law patriotism. It life. Still I fight you. David Cole. The pro-life **pro-second Amendment** conservative choice state senate. | Democrat Terry Goodman office, it's time change change, Southern Indiana, bringing high paying jobs Indiana work Republicans Tuesday, November 3rd. I'm asking phone. |
| Actual = 1 | Gunfire Chaos Texas. Running panic police scrambling police shielded gunman. Lost livesbetter **getting guns streets**. We **protect** Texan. | We want tell Eddie Rodriguez, growing Rio, Grande Valley. He learned fight families. Austin. **He fought NRA**. He fought free School breakfast Texas. Child. He fought healthcare crisis. She fight Texas. We know family. |

*Figure 5: Examples of results of our models predicted gun control position compared to actual gun control position with transcripts with stopwords removed.*

When we look at some examples of how our model predicts the sentiment/tone of an ad, we can see in the examples of a true positive and a true negative that there are specific phrases that it seems like the model is recognizing, and from those specific phrases determining if the overall tone is positive or negative around gun control. In examining the data, we saw for anti-gun control common phrases include “defend the second amendment,” “pro-gun,” “protect gun rights” and things like that, while for pro-gun control, common phrases included things like “increased background checks,” “fight gun violence,” “gun safety laws.” In the example of the false positive above, we see that in the transcript there isn’t even a mention of anything referencing gun control, even though it’s been human coded as anti gun control. This could be due to the fact that in the ad there may be text on the screen, or simply an image of someone carrying a gun which could show that the candidate is anti gun control without words in the transcript reflecting this. In the example of a false negative, when we remove stop words we just have the phrase “getting guns streets” however with the stopwords, the original phrase was “ getting guns off our streets” which gives this phrase a very different meaning, and in this case we also have the word “protect” which is very common in anti gun control ads with the phrase “protect the second amendment” so it might also be picking up the word protect in the wrong context, both of which might contribute to make it difficult for the model to predict the tone.

|  | Average (from tuning or cross-validation) | Maximum (from tuning or cross-validation) |
| --- | --- | --- |
| Logistic Regression | 0.886 | 0.92 |
| Naive Bayes | 0.870 | 0.87 |
| Random Forest | 0.878 | 0.89 |
| SVM | 0.856 | 0.89 |

*Figure 7: Accuracy results of logistic regression with cross validation, naive bayes with no changes (so no difference in average and maximum), and random forest and SVM with tuned hyperparameters.*

We found that the logistic regression model performed the best out of the four algorithms by comparing the average accuracy from the cross-validation or tuning of the dataset. Accuracy depicts the total number of correct predictions divided by the total number of predictions made, and taking the average accuracy decreases the chance of error based on a skewed/biased training set. In our case, on average the logistic regression model for our subsetted data correctly identified the gun control position of the ad 88.6% of the time, while the other models were only slightly below that. We compared both the average accuracy of the model (over either cross-validation of different splits of the data or with different values for the parameters) and the maximum accuracy of the model (over the same options for different accuracies) to see how well the model would predict the data in the optimal case.

Additionally, we wanted to see how our models performed using labeled training data compared to a pre-trained model. In comparing these results to VADER, on the entire dataset, the VADER model only had 36.4% accuracy. VADER may have performed poorly because it was trained on social media data, which provides a different environment for the usage of some words. For example, candidates will sometimes use negative-connotated words to enforce a point that they are pro-gun laws. If a model is not trained to recognize this word usage it may confuse the results. When we subsetted the data to include gun imagery, we did not see much change in the results of the VADER model with an accuracy of 36.6%.

*VADER Accuracy on entire dataset = 0.364; Accuracy for gun imagery subset = 0.366*

We also wanted to see how our model’s accuracy might change if we subsetted our data to only include ads that included a gun image in them. One thought we had going into this project was that the presence of a gun in the ad might mean that it would be easier to predict the sentiment of the ad, as a gun is an intense image to include, and it might reflect in the words of the ad also being more intense and therefore easier to classify. However, after running the logistic regression, we found that the subset of gun images only had a very slightly better classification result. The average accuracy of the 10-fold cross validation was 88.8% compared to our 88.6% average accuracy on the whole dataset.

*Accuracy results of logistic regression with subset of ads containing gun images =* ***0.888***

|  |  |
| --- | --- |

*Figure 6: Data subset breakdown based on pro or anti gun control.*

As shown in Figure 6, of the 500 videos, we saw that 28% of the advertisements that discussed gun control were pro-gun control and 71% of the ads that discussed gun control were anti-gun control. When we examine how candidates use gun depiction, it was eye-opening to analyze the use of gun images, as 28% of political ads discussing gun laws utilized gun imagery. 75% of ads discussing gun laws were anti-gun control. It is interesting to have a breakdown of the use of gun imagery in tandem with gun control topics, as it shows how guns are being used in campaign ads to sway the general public. Our dataset would suggest that guns are used significantly more often when a candidate is anti-gun control, while pro-gun control candidates’ ads steer away from using gun images.

The 2018 Barry study revealed an increase in the mention of guns in candidate-related TV airings over the previous four election cycles. While pro-gun rights ads dropped 41% from 2012 to 2018, the share of the ads referring to guns increased from 1% to 8% over the same time period. Barry hypothesized that the insights and trends from that study may show a developing trend for future election cycles. From our dataset, only 5.4% of the entire dataset was identified as discussing gun control at least once in the ad. An even smaller percentage of the dataset took a stance on pro or anti-gun control, as a proportion of ads labeled as mentioning gun control were labeled as miscellaneous discussion of gun control (meaning they took no official stance). While this statistic is slightly different from any mention of guns, these two topics typically go hand-in-hand. Furthermore, our dataset was much smaller than Barry’s, which could influence the proportions of different ad topics. However, our data shows a smaller percentage of gun control mentions than we would have predicted given Barry’s study results and their future hypotheses.

# Conclusion

With a goal to investigate how candidates in the 2020 election cycle discuss gun laws, we trained a sentiment classifier on gun control political ads to create a model that can be used in future election cycles to identify subject matter trends. For example, given an unlabeled political ad (ad without human coded values), our image classification model could be used to identify cases of gun images, which could then be used in combination with the sentiment classifier to identify gun control discussion. Alternatively, the sentiment classifier could be used alone to determine gun control position without image analysis.

Using the TV advertisement sponsored by federal candidates in the 2020 general election, we focused on the detection of gun images and mention of gun control in political television ads. We made a subset of the dataset to only include ads that discussed gun control, and utilized object detection and image classification algorithms with key frames from the ads to identify gun images using the RegNet\_y\_128f model. This allowed us to create another subset of the data that had gun imagery in it. Creating these subsets allowed us to apply our sentiment classifier to investigate different features from the data later on. Front there, we began sentiment analysis. Using the same dataset of 2020 TV ads, we merged it with the transcript data of those ads, generated from Google’s Speech-to-text. We only selected ads with ads whose transcript confidence was greater than 0.8, and had been human-coded for issue\_proguncontrol or issue\_antiguncontrol. This was important, as this allowed us to analyze word usage and connotation with human-coded variables like pro and anti gun control, which illustrates the candidates view on gun laws. These two qualifications left us with 500 entries. The image classification data from earlier was used to further subset the data to only include gun images, shrinking the subset to 142 entries. Performing an 80/20 split, we used Logistic Regression, Naive Bayes, Random Forest, and SVM models with cross-validation or tuning on the larger dataset. After comparing the accuracy of the models, logistic regression was the best performer.

Our final sentiment analysis model using logistic regression performed with 88.6% accuracy on predicting whether a candidate is pro or anti gun control. When subsetting the data further to include ads that have guns depicted we saw a slight increase in accuracy, as the subset of ads with guns depicted was 88.8%, and depending on the split of the data sometimes had greater than 95% accuracy. This may indicate that ads with gun imagery contain more words with strong connotations, whether they be negative or positive, as the model performed slightly better with that subset. When using the VADER model, we saw a dramatic decrease with an accuracy of 36.4% while when we subsetted to include gun depictions we got an accuracy of 36.6%.

To the broader research topic involved in identifying trends in political ads, our classifier could be used as a pre-trained model for a specific subgenre of American politics. It can be used with unlabeled data in order to gain information about the aim of the ads of future election cycles. Having been trained on political ads featuring discussions of gun control, it would be a good classifier to use on political ads in the future to see prospective trends involving gun usage and gun laws.

For future research purposes, it would be interesting to train a sentiment analysis model on the entire dataset instead of just the videos that discuss gun control. This could be used to investigate political advertisement jargon and how to classify it. Furthermore, we could perform sentiment analysis including the background music of the ads as well as their transcripts. Another potential future research question could be centered around sentiment analysis of gun control topics. This would be focused on aspect-based sentiment analysis on the content of the ads. Moreover, this could be experimented with different types of video content, such as YouTube videos or social media videos to see public sentiment. On the image side, we would be interested in training our own image detection model using the Wesleyan Media Project dataset, where there could be a greater range of images included in training. Specifically, we would want to include image classification with NRA symbols or other symbols that represent gun usage. This would allow for a greater range of images to be classified in the subset of ads with the gun control topic and gun-related imagery.

# 

# References

Barry, C. L., Kennedy-Hendricks, A., Bandara, S., Baum, L., Gollust, S. E., & Niederdeppe, J. (2020, February). Guns In Political Advertising Over Four US Election Cycles, 2012–18. *Health Affairs*, *39*(2). https://doi.org/10.1377/hlthaff.2019.01102

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 10012-10022).

Malde, Ravi. “A Short Introduction to VADER.” *Medium*, Towards Data Science, 8 July 2020, https://towardsdatascience.com/an-short-introduction-to-vader-3f3860208d53.

Radosavovic, I., Kosaraju, R. P., Girshick, R., He, K., & Dollár, P. (2020). Designing network design spaces. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10428-10436).

Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K. (2017). Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1492-1500).