

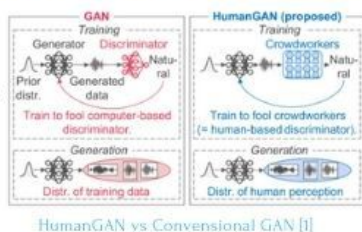
How can crowdsourced workers effectively rate artwork images produced by Generative Adversarial Network transformers?

Moshiur Rahman
(M.Rahman-4@student.tudelft.nl)

Responsible Professor(s):
Derek Lomas,
Ujwal Gadiraju
Supervisors:
Willem van der Maden & Garrett Allen

(1) BACKGROUND INFORMATION

- GATHER HUMAN RATINGS TO FEED BACK TO THE GAN GENERATORS TO PRODUCE MORE AESTHETIC IMAGES. THIS TECHNIQUE IS CALLED HUMANGAN
- HUMANGANS ARE MORE EFFECTIVE THAN CONVENTIONAL GANS WHEN IT COMES TO EVALUATING/DISCRIMINATING GAN-PRODUCED HUMAN EXPERIENCES (ARTWORKS OF LANDSCAPES IN THIS CASE)[1]
- INVESTIGATE BETWEEN THE BINARY-CHOICE FORMAT AND THE FOUR-CHOICE FORMAT



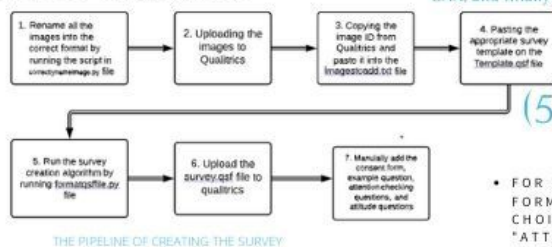
Examples of some of the images used in the experiment. First row is from TCDNE GAN, then the Satellite GAN, then the Baseline GAN, and finally the GAN250 GAN

(2) RESEARCH QUESTIONS

MAIN QUESTION: HOW CAN CROWDSOURCED WORKERS EFFECTIVELY RATE ARTWORK IMAGES PRODUCED BY GENERATIVE ADVERSARIAL NETWORK TRANSFORMERS?

(3) METHODOLOGY

1. AUTOMATE THE CREATION OF THE SURVEY. TIME DECREASED FROM 1 H 30 MINS TO 10MINS
2. ADD EXTRA SURVEY FEATURES
 - A. CONSENT FORM
 - B. EXAMPLE QUESTIONS
 - C. ATTENTION CHECKING QUESTIONS
 - D. ATTITUDE QUESTIONS

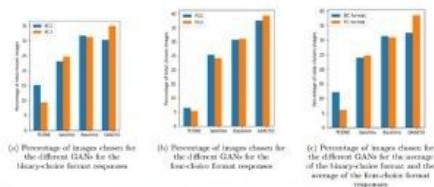


(5) CONCLUSION

- FOR EVERY TEST, THE FOUR-CHOICE FORMAT WAS BETTER THAN THE BINARY-CHOICE FORMAT, EXCEPT FOR THE "ATTITUDE OF THE PARTICIPANTS" TEST.
- THE RESULTS WERE MORE CONSISTENT AND RELIABLE FOR THE FOUR-CHOICE FORMAT
- IMAGE-GROUPINGS HAD FEWER EFFECTS ON THE FINAL RESULTS FOR THE FOUR-CHOICE FORMAT

(4) RESULTS

1. REMOVING POOR RESPONSES.
 - A. REMOVE ALL RESPONSES WHERE MORE THAN ONE ATTENTION-CHECKING QUESTIONS (ACQ) WERE WRONGLY ANSWERED.
 - B. CONDUCT A CHI-SQUARED TEST OF INDEPENDENCE ANALYSIS BETWEEN RESPONSES WITH ONLY ZERO AND ONLY ONE ACQ MISTAKE. AFTERWARD, ANOTHER TEST BETWEEN ZERO OR ONE ACQ MISTAKE(S) AND ONLY ZERO MISTAKES
2. GOODNESS OF FIT TEST
 - A. CONDUCT A CHI-SQUARED TEST BETWEEN THE IMAGE RATINGS AND THE EXPECTED VALUE
3. EFFECTS OF IMAGE GROUPINGS ON THE SURVEY RESULTS
 - A. CHI-SQUARED TEST OF INDEPENDENCE BETWEEN THE RESULTS OF THE TWO SURVEYS FOR BOTH FORMATS
 - B. CALCULATE THE JENSEN-SHANNON DIVERGENCE DISTANCE BETWEEN RESULTS
 - C. CALCULATE THE PEARSON CORRELATION COEFFICIENT BETWEEN RESULTS
4. ATTITUDE OF PARTICIPANTS
 - A. ANALYZE THE ATTITUDE OF THE PARTICIPANTS BETWEEN THE BINARY-CHOICE AND THE FOUR-CHOICE FORMAT



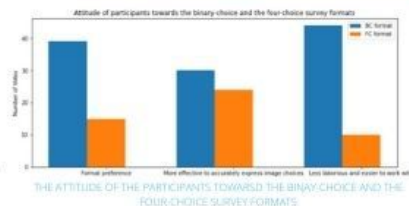
SURVEY RESULTS FOR THE TWO FORMATS

GAN name	FC1	FC1 shown	FC2	FC2 shown	BC1	BC1 shown	BC2	BC2 shown
TCDNE	115	2368	63	1152	310	1024	230	1250
Satellite	508	2368	278	1152	472	1024	631	1250
Baseline	723	2368	358	1152	649	1024	790	1250
GAN250	802	2368	453	1152	617	1024	891	1250

RESULTS AFTER REMOVING THE POOR RESPONSES

p-value	BC1 vs BC2	FC1 vs FC2	Average of BC surveys vs average of FC surveys
	<0.001	0.480	<0.001

THE P-VALUES FOR THE DIFFERENT SURVEYS OF THE DIFFERENT FORMATS



(6) FUTURE WORK

- GROUP LANDSCAPE IMAGES ACCORDING TO BIOMES. THIS CAN HELP IDENTIFY WHETHER SOME GANS CREATE BETTER IMAGES FOR ONE BIOME BUT WORSE FOR ANOTHER
- MORE EXTENSIVE CONSISTENCY TEST FOR PARTICIPANTS TO HAVE MORE RELIABLE RESULTS
- CREATE MORE VERSIONS OF EACH SURVEY FORMAT. LESS EFFECTS OF IMAGE GROUPINGS