Loss

1. חילוץ פיצ'רים בעזרת CNN נעשה בעזרת חילוץ מאפיינים מרחביים דומים בתמונות. השונות הנוצרת כתוצאה מהשוואה זו יכולה לשמש כפונקציית מחיר הלוקחת בחשבון פרמטרים מרחביים הקשורים לאופי התמונה, ולאפשר אימון רשת לסופר רזולוציה. מכיוון שפונקציות מחיר רגילות משתמשות בנורמות, הן מקטינות הזיות פרטים קטנים ברזולוציות הנמוכות אך יוצרות טשטוש. שימוש ברשת, מאפשר ללמוד את פילוגי התמונות, ובכך להשלים יותר טקסטורות – במחיר של פרטים לא הגיוניים (שכן לא כל התמונות מתאימות מספיק זו לזו). שימוש במחיר זה יוצר תמונות ריאליסטיות יותר.
2. שחזור תמונה התקדם בשנים האחרונות תודות לרשתות עמוקות. אבל, כאשר מאמנים רשתות עמוקות, מידת הריאליסטיות של תמונות טבעיות נפגעת.
3. פונקציות המחיר הנפוצות הן נורמות (1\2), אך הן סובלות מבעיית התכנסות לממוצע.
4. רשתות גנרטיביות מצליחות ליצור שחזורים יותר מציאותיים, אך קשה לאמן אותן (לדוגמה, GAN שאינו מתכנס בקלות).
5. בנוסף, GAN מתאים את עצמו למרחב התמונות הנפרשות מסט האימון, ולכן ישנן הזיות.
6. מחקרים אחרים (15) הראו שרשתות לזיהוי תמונה מכילים מידע על המאפיינים העמוקים (מרחביים) של התמונות.
7. במאמר זה, משתמשים הפונקציות מחיר חדשה כדי להקטין את בעיית ההתכנסות לממוצע, וניתן לראות כי כמות ההזיות קטנה.
8. גישה זו דומה יותר לרשת neural style transfer, שבו אנו משתמשים בסטייל של תמונה מנחה אחת לתמונה אחרת. המחיר של השוני בתוכן מוגדר ע"י המרחק באלמנטים של הפיצ'רים.
9. נראה כי המאפיינים שחילצנו שייכים למרחב מטרי, ולכן ניתן להגדיר גיאומטרי עם מרחקים, ואז להגדיר בעיית optimal transport למרחקי W.
10. אנו משתמשים בsliced Wasserstein (שהוגדר ב26) שלו קיים פתרון סגור וקל לחישוב. אנו מחשבים אותו במימד נמוך (כל פועם מוטל רנדומלית על מחרב אחר) ומבצעים התמרת רדון על מנת לשחזר\לשערך את Wasserstein הכללי.
11. המסגרת שלנו לניתוח התמונה משתמשת ב-VGG על מנת לחלץ את המאפיינים מתמונת המקור (ground truth), בעזרת הפיצ'רים מה-VGG מבצעים השוואה בין המקור לשחזור, ומודדים בעזרת W. הגישה נקראת PDL, ואינה מוסיפה סיבוכיות חישובית.
12. האלגוריתם נבחן בסופר רזולוציה, הסרת רעש ושחזור רעשים מ-JPEG, ומספק תוצאות טובות יותר מרחבית ללא הוספת עיוותים משמעותיים.
13. שימוש ב-CNN בעזרת נורמת L2, היו טובות בהשוואת מודלים רדודים או להסיק על פילוג סט האימון. נעשה שימוש בהגדרת loss לפיקסלים.
14. מאמר אחר, הדגים את השימוש במאפיינים עמוקים כדי חתפוס מידע מרחבי. שימוש במאפיינים עמוקים אלו משמש כדי למצוא דימיון מרחבי באופן מוצלח יחסית. נראה שפרמטרים אלה קורלטיבים עם התפיסה האנושית של מאפיינים בתמונות, ולכן שילוב של פרמטרים אלה במודלים מביא לתוצאות טובות יותר.
15. כאשר מנסים לשפר תמונות טבעיות CTX נתן תוצאות לא רעות, אך כאשר הוגדר CoBI שלוקח בחשבון יותר מרחקים בין פיצ'רים באיזורים קרובים (כלומר, מושפע יותר מרחבית) מקבלים שחזור טוב יותר (הערה: פחות מוצלח לגנרטיביות כללית – כי לא תמיד אנחנו רוצים תלות מקומית).
16. שימוש ב-GAN נהיה מקובל מאוד בבעיות אלה, בה רשת אחת מייצרת תמונות, ורשת שנייה מאומנת על מנת למצוא את הזיוף מהמקור. לפעמים משלמים בעיוותים בשחזור תמונות ברזולוציה נמוכה.
17. במאמר מוצע PDL, שבו המחיר משולם על הבדל במאפיינים העמוקים בין המטרה לשחזור. השוואה נעשית מקומית **וגם** בעזרת התפלגויות המאפיינים, וכך ניתן לחלץ תמונות ריאלסטיות עם פחות הזיות בהשוואה ל-GAN. מתבצע שימוש בSW על מנת להשוות התפלגויות חד מימדיות של היטלי המאפיינים, כך שהוא משלים את הבדלי הפיקסלים המקומיים.
18. ווסרשטיין משתמש בגיאומטריה של המרחב כדי 'להעניש' על הבדלים בין ההתפלגויות. בדוגמאות שצורפו (גרף 2) ניתן לראות איך ווסרשטיין רגיש יותר לשינויים בצורת ההתפלגות. מכיוון שהתפלגויות הן מרחב מטרי (לבדוק למה).
19. למדד ווסרשטיין אין פתרון סגור, והוא בודק את הקשר בין פילוג ההסתברויות הרב ממדיות.
20. ווסרשטיין חד מימדי מאפשר שימוש בCDF של ההתפלגויות על מנת לקבל תוצאה, ולכן ביישום נומרי, ניתן לסכום בעזרת גבולות האינדקטורים המגדירים את ה-CDF. לכן, בחד מימד, ניתן להעריך בקלות את מדד מרחק ווסרשטיין.
21. SW משתמש בהטלה כדי לחשב חד מימדית, ואז נעשה שימוש בהתמרת רדון כדי לקבל קירוב רב מימדי. משתמשים במונטה קרלו כדי לשערך את האינטרגל בביטוי. כמו הדגימות שצריך כדי לשערך גדלה אקספוננטציאלית במספר הדגימות. רוב הכיוונים אינם משמעותיים, ויש מחקרים הבודקים איזה כיוונים כדאי לבדוק.
22. ה-PDL מוגדר ע"י חילוץ הפיצ'רים מרשת VGG שאומנה כבר, אותם אנו מטילים ע"פ מטריצה W (איך בדיוק לבחור – בעתיד ימצאו דרך חכמה יותר, וכרגע רנדומלי). המודל במאמר מניח כי הפיצ'רים בלתי תלויים – כך ניתן לקרב את מרחק ווסרמן באמצעות ביצוע מיון לערכי הפיצ'רים, ומציאת המרחק ביניהם בעזרת נורמה רגילה. במידה והייתה ביניהם תלות, לא היינו יכולים לבצע את הקירוב הזה, שכן חיסור הפיצ'רים לא יבטא את המרחק, ונאלץ לפתור את האינטגרל.
23. ה-PDL שלנו מוגדר ע"י נורמת q בין התוצר למטרה בתוספת ממושקלת של הקירוב שביצענו לווסרשטיין. ההתפלגות המוצעת לוקחת בחשבון תזוזה מרחבית של הפיצ'רים – גמישות זו מאפשרת הימנעות מבעיית ההתכנסות לממוצע.
24. ההבדל בין Perceptual לPDL הוא בהטלת הפיצ'רים (משוואה 6, יש את שאלה 8 בנושא, כי לא ברור). התקבלו תוצאות מטושטשות.
25. Contextual Loss דומה מאוד לPDL מחיר זה נקבע ע"י הדמיון בין הפי'צרים המתאימים בתמונות (יש הסבר למדדים, פחות מבין אותם). ווסרשטיין מספק יתרון בשיפור תמונות בהשוואה.
26. בכל הניסויים, ניסו למצוא אץ האיזון הטוב ביותר בין התאמת הפיקסלים למשקל ההתפלגויות המרחבי. בעיה זו הינה קשה לחישוב..
27. במאמר השתמשו בארכיטקטורות שונות של CNN למטרות שונות (ניקוי הרעשה והסרת JPEG, סופר רזולוציה, והורדת טשטוש).
28. הערכת ההפסדים המרחביים נעשתה בעזרת שקלול של שלושה מדדים, המשתמשים לדימיון. **לא בטוח שאנחנו נצטרך את אותן המדדים.**
29. הוספת רעש חזק לתמונות, וניסיון הסרת הרעש בעזרת ה-PDL וגישות מקובלות אחרות. נעשו מספר ניסיונות אמפירייים למציאת המשקל האופטימלי בPDL. לבסוף קיבלו תוצאות טובות יותר במודל המוצע של PDL. יש טריידאוף בין הנראות לבין ההזויות ועיוותים.
30. באופן כללי, הPDL מספק תוצאות טובות, כאשר contextual loss הכי קרוב אליו מבחינת ביצועים.
31. בסופר רזולוציה, SRGAN יוצר מעט יותר רעש 'גרגרי', אך נוטה לפחות 'הזיות'. באופן כללי, ה-PDL ו-SRGAN מספקים תוצאות טובות יותר ביחס לגישות אחרות.
32. הורדת טשטוש –
33. נראה כי אמפירית, אין הבדל בין השוואת פי'צרים באופן בלתי תלוי לבין הטלות רנדומליות. בנוסף, הפיצ'רים אינם מנורמלים מה-VGG (ערכים שונים וטווחים שונים), ונרמול שלהם הביא לעיוותים – כלומר, יש פיצ'רים שווים ויש פיצ'רים ששווים יותר. זהו פתח לעבודה עתידית.
34. לסיכום, נראה שהחד מימדיות מאפשר הגעה לתוצאות טובות, ובעתיד ירצו למצוא הורדת מימד טובה ובת"ס באמת.

שאלות:

1. עמוד 1 – הסתברויות מקומיות – איך מודדים כאן משהו מקומי? הכל נראה נקודתי או גלובלי.
2. מחרב מטרי – למה המחיר הוא מרחב מטרי? מה גורם לנו לשייך לשם גם הסתברויות? האם התועלת היחידה היא הגדרת גיאומטריה ושימוש בoptimal transport?
3. על מה ה-VGG אומן? מאיפה לוקחים את הרשת? איך זה משפיע על הפיצ'רים.
4. האם אנחנו נרצה לקחת בחשבון גם הבדלים בין ערכי פיקסלים בפונקציית המחיר שלנו? לכאורה זה מקומי מדי ואנו נרצה כללי. אולי משקלים שונים?
5. לעבור על נוסחה 4 – SW , לא לגמרי ברור לנו.
6. הפתרון הסופי נראה בלתי תלוי ברדון \ מונטה קרלו וכו' בגלל הנחת אי התלות. האם פספסנו משהו?
7. על נקודה 24 – איך בוחרים את ההטלה? מה גורם להטלה אופטימלית? אם זה כל ההבדל, גם הטלת היחידה היא הטלה, מדוע הביצועים נפגעים?
8. האם המדדים שמצויינים במשוואה 7 רלוונטים גם ליישום שלנו?

תשובות של תמר:

1. זה local כי בסוף הוא מסתכל על פיצ'רים מתוך התמונה. כל פיצ'ר במפת פי'צרים היא רלוונטית לאזור מסוים בתמונה. כאן מסתכלים על התפלגות של הרבה תמונות.
2. מרחב מטרי – שאלה טובה (?) יש מצב שזו פשוט הנחה. כנראה לא קריטי.
3. ה-VGG זו רשת מאומנת על קלסיפקציה.
4. אצלנו הפיקסלים ממש פחות מעניינים, אנחנו רוצים אצלנו לעשות גנרציה
5. בהמשך אם יהיה רלוונטי.
6. בגלל ההנחה על ב"ת אפשר לזרוק את הפיתוח כדי להצדיק את WS1D.
7. לא בטוח האם באמת עשו הטלה בניסוי. כנראה הם עשו בלי הטלה (נאיבי)
8. הscore לא מקובל, ביקשו להוריד את זה. בגרסה הסופית כנראה שזה ירד.