**האוניברסיטה הפתוחה**

**המחלקה למתמטיקה ולמדעי המחשב**

**שיפור רזולוציה של תמונות בעזרת רשתות GAN**

פרויקט גמר זה הוגש כחלק מהדרישות לקבלת תואר

"מוסמך למדעים" M.Sc. במדעי המחשב

באוניברסיטה הפתוחה

המחלקה למתמטיקה ולמדעי המחשב

על-ידי

**ליאור בומוורצל**

הפרויקט הוכן בהדכרתה של ד"ר מיריי אביגל

פברואר 2023

# תקציר

הגדלת רזולוציית התמונה משמעה הפיכת התמונה לחדה וברורה יותר, על ידי הוספת פיקסלים נוספים לתמונה. בפרויקטים כאלה, אלגוריתמים המבוססים על רשתות יריבות זוכים לפופולריות רבה כיוון שהם מייצרים תמונות שנראות אמיתיות וחדות, ויש להן יכולת השלמת טקסטורות הנלמדות מבסיס הנתונים. בשלב האימון, האלגוריתמים מתאמנים על מאגרים רבים של תמונות ברזולוציה גבוהה ולומדים את התכונות והטקסטורות השוליות המאפיינות את העולם האמיתי. במהלך שלב ההפעלה, האלגוריתמים משחזרים את הטקסטורות הללו ומשלימים את המידע הנדרש בהתאם לדרישות המשתמש. בעקבות השלמת הטקסטורות במקומות בהם חסר מידע בתומנה, נוצרת תמונה ברזולוציה גבוהה יותר ובאיכות טובה יותר. אלגורתמים המשתמשים בארכיטקטורה של רשתות יריבות, מייצרים תמונות איכותיות אך לא בהכרח התמונות הכי קרובות לתמונות המקור שרוצים לשחזר, הסיבה לכך היא שהאלגוריתמים מתמקדים בשיפור איכות התמונה על חשבון דיוק השחזור.

תוכן עניינים

[תקציר 2](#_Toc140586196)

[1.מבוא ...................................................................................................................................4](#_Toc140586197)

[2.סביבת העבודה................ 5](#_Toc140586198)

[3.ייצור הדאטה........... 5](#_Toc140586199)

[4.הסבר על מבנה הפרויקט........................................................................................................................6](#_Toc140586200)

[5.הסבר על אופן מדידה והשוואה בין הארכיטקטורות......................................................................... 6](#_Toc140586201)

[6.SrResNet 8](#_Toc140586202)

[SrGan.7 11](#_Toc140586203)

[8.בעיית קריסת מצב.................................................................................................................13](#_Toc140586204)

[9.WGAN 17](#_Toc140586205)

[10.WGAN-GP 19](#_Toc140586206)

[11.Pix2Pix 21](#_Toc140586207)

[12.SinGan 24](#_Toc140586208)

[13.מציאת התמונות עם השיפור הגדול ביותר בין הארכיטקטורות. 25](#_Toc140586209)

[14.הצגת התפלגות ווקטור המפאיינים ברשת inception 26](#_Toc140586210)

[15.סיכום............ 27](#_Toc140586211)

[16.רשימת מקורות 28](#_Toc140586212)

## מבוא

הגדלת רזולוציה בתמונות דורשת לייצר מידע אשר לפעמים לא ניתן לשחזר מהתמונה הקטנה מכיוון שמידע מתמונה מוגדלת יכול להיות מוצג ממספר גדול של פיקסלים ובתמונה המוקטנת יוצג כפיקסלים בודדים. הטקסטורה אשר אוחדה לכמה פיקסלים בודדים, צריכה להתרחב למספר פיקסלים יותר גדול, ועם יותר הפרדה.

בעבודה זו, אבצע שימוש באימון רשתות יריבות בארכיטקטורת, Adversarial Generative Networks [1] בראשי תיבות GAN לצורך הגדלת רזולוציית תמונה. ארכיטקטורות GAN מאפשרות למידה של התפלגות המידע מהעולם האמיתי ושימושו בשעת יצירת התמונה המוגדלת בהשלמת המידע החסר.

מאז מאמר [1] בו ארכיטקטורת GAN הודגמו לראשונה, התפרסמו עוד מגוון רחב של ארכיטקטורות אשר בכולן העקרון הבסיסי , שימוש ברשתות יריבות המתחרות ביניהן. מטרת האחת (הרשת הגנרטיבית) הוא ליצור תמונה, ומטרת השנייה (הרשת הדיסקרימנטיבית) לסווג האם התמונה באה מהעולם האמיתי או יוצרה על ידי הרשת הגנרטיבית.

לצורך מימוש הגדלת רזולוציה, בעבודה זו אתבסס ראשית על מאמר [2] אשר הציע את ארכיטקטורתSRGAN , אותה נתאר בהרחבה בהמשך. המאמר הראה הגדלת רזולוציה עד פי 4 מתמונת מקור מוקטנת. לאחר שאראה את יכולות הארכיטקטורוה על מספר דוגמאות, אשתמש בעקרונות אשר הוצעו במאמרים הנוספים לארכיטקטורות GAN לצורך בחינה האם הן יכולות להוות שיפור לארכיטקטורת SRGAN.

לצורך השוואה בין תוצרי הארכיטקטורות אשתמש בכמה מדדים להשוואה בין תמונות, אשר ארחיב עליהם בהמשך. המדדים יבדקו את התוצרים בכמה מובנים של מרחק בין תמונת מקור לתמונה המיוצרת, כגון מרחק פיקסלים בין תמונת מקור לתמונת מוצא, או השוואה של המאפיינים והסטטיטיוקות של מקבצי תוצרי הרשת הגנרטיבית למול מקבצי תמונות אמיתיות.

עבודה זו היא המשך של עבודה מסכמת אשר ביצעתי בנושא שיפור איכות תמונה בעזרת רשתות GAN. לצורך השלמת הסברים אפנה לעבדוה המסכמת.

## סביבת העבודה

הפרויקט נכתב בשפת Python3 בעזרת שימוש TensorFlow גרסה 2.10.  
 TensorFlow היא ספריית קוד פתוח ללמידת מכונה, המפותחת על ידי חברת גוגל לבנייה ואימון רשתות עצביות.

ספריית TensorFlow נותנת פתרונות ותשתית לאימון מודלים ללמידת מכונה אשר בהם השתמשתי:

* TensorBoard משמש כסט כלים לוויזואליזציה של TensorFlow. הוא מספק ממשק משתמש ידידותי בו ניתן לבחון ולפרש את הריצות והגרפים של התקדמות האימון לגרפים השונים.
* Keras הוא API של רשתות עמוקות, כתוב ב-Python ומסוגל לרוץ על TensorFlow כ-API ידידותי למשתמש. Keras מקל על בניית ואימון מודלים של למידה עמוקה. הוא מספק פונקציונליות להגדרה ואימון מודלים של רשתות עמוקות .Keras גם תומך במגוון רחב של סוגי שכבות, פונקציות הפסד, אופטימיזטורים, ועוד.
* ניתן לטעון נתונים ולעבדם מראש באופן יעיל באמצעות API tf.data, שיכול להתמודד עם כמויות גדולות של נתונים, לבצע המרות מורכבות ולטעון נתונים באופן מקביל ויעיל.

החומרה עליה הורצו המודלים והבדיקות :

מחשב 9500 XPS Dell

מעבד Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz 2.59 GHz

זיכרון RAM 32.0 GB (31.8 GB usable)

GPU NVIDIA GeForce GTX 1650 Ti

## ייצור הדאטה

בעבודה זו השתמשתי בדאטהסט CelebFaces Attributes (CelebA) Dataset

בפרט הורדתי את img\_align\_celeba מהאתר של Kaggle. מקבץ התמונות מכיל כ 202.5 אלף תמונות של פרצופי מפורסמים, התמונות מיושרות וחתוכות כך שדמות המפורסם במרכז התמונה. גודלי התמונות המקוריות הן 218X178.

לטעינת התמונות ייצרתי מחלקה עוטפת הנקראת MyDataLoader הנמצאת בקובץ DataLoader. המחלקה מייצרת מקבצים (באטצ'ים) של טנסורים בעזרת שימוש בתשתית הניתנת על ידי ה tf.data API של TensorFlow. המחלקה מקבלת כפרמטר את גדלי המקבצים( באצטצ'ים), גודל התמונה הגדולה, יחס ההקטנה, כמות התמונות אשר יישמרו למקבץ המבחן .validation תמונות ממקבץ ה validation לא יוצגו למודלים בשלב האימון, תמונות אלה ישמשו לבדיקת איכות תוצרי הארכיטקטורות השונות.

בעבודה זו השתמשתי בפרמטרים לאימון הדומים לפרמטרים לאימון מהמאמר [2]. גודל התמונה ברזולוציה הגבוה 96, יחס ההקטנה לרזולוציה הנמוכה 4. השארתי 5% ממקבץ התמנות ל validation, זוהי כמות גבוהה יחסית של מקבץ validation. הסיבה לגודל הנבחר בסט הוולידציה היא שביצעתי שימוש בתמונות הוולידציה במדידת מטריקות למרחקים בין מקבצי תמונה אשר דורש כמות גדולה יחסית של תמונות לצורך תוצאה מדויקת יותר. בשלב טעינת התמונות אני מבצע צימוד של כל תמונה ברזולוציה גבוה לתמונה אשר עוברת הקטנת bicubic, כך שכל איבר מהדאהסט שאני יוצר מורכב מ-2 תמונות - אחת ברזולוציה גבוהה והשנייה ברזולוציה נמוכה.

## הסבר על מבנה הפרויקט.

הפרויקט נועד לעשות השוואה בין ארכיטקטורת GAN . בחרתי להשתמש בעקרון של פולימופיזם ולממש מחלקה אבסרטקטית כך שהקוד עבור הפעולות המשתפות ייכתב פעם אחת. תפקיד המחלקה האבסטרקטית הוא להוות בסיס משותף לאימון. לולאת האימון, מדידות הביצועים, שמירת תוצרי הביניים וחישוב המדדים לאורך התקדמות האימון נעשים במחלקה זו. בעקבות מימוש מחלקה זו, עבור כל מודל שאבדוק בפרויקט, הדבר היחידי שצריך לעשות על מנת להעריך את טיבו הוא לרשת מהמחלקה האבסטרקטית ולממש את צעדי האימון לגנרטור והדסקרימנטור. הארכיטקטורות השונות אשר אני בוחן, עושות את השינוי הנדרש בקבצים שונים שקרויים לפי שם הארכיטקטורה כפי שאסביר במהלך העבודה. המחלקה האבסטרקטית נמצאת בקובץ train.py ונקראת AbsTrainer.

לשם הפרדה בין צורת האימון לייצירת המודל, בחרתי להפריד את יצירת המודלים לקובץ אשר מיועד לכך ונקרא model.py. בקובץ זה, אני בונה את כל המודלים השונים המתאימים לארכיטקטורות השונות.

יצירת מחלקה display\_handler שמממשת את הצגת התמונות ושמירתם, המחלקה ממשת הצגה של גרפים לפיצרים ומציאת אזורים בהם חלקי התמונה שונים ביותר.

בקובץ run.py אני טוען את התמונות לאימון ולבחינה, מייצר את האובייקטים המממשים את המחלקות השונות ומריץ את האימון שלהם אשר מייצר את התוצרים של הרשת הגנרטיבית, מודד את ערכי המדדים השונים, ושומר את המשקולות לאורך התקדמות האימון.

בקובץ model.py אני ממש את המודלים השונים, בקובץ זה על פי פרמטרים נבחרים השכבות, כמות וגדלי הפילטרים.

## הסבר על אופן מדידה והשוואה בין הארכיטקטורות.

אלגוריתמים המשתמשים בארכיטקטורה של רשתות יריבות, מייצרים תמונות איכותיות אך לא בהכרח התמונות הכי קרובות לתמונות המקור שרוצים לשחזר, הסיבה לכך היא שהאלגוריתמים מתמקדים בשיפור איכות התמונה על חשבון דיוק השחזור. לצורך שיפור איכות התמונה האלגוריתמים מוסיפים מידע וטקסטורות אשר למדו בשלב האימון מאובייקטים מהעולם האמיתי.

אין הגדרה מסויימת איך התמונות הסופיות בתוצרי הרשת אמורים להראות. האם ייצור תמונה עם מאפיינים חדים אך שונים במיקצת מהמקור עדיף על תמונה מטושטשת במיקצת אבל עם מאפיינים דומים למקור?

השוואה בין ארכיטקטורת לרשתות יריבות זו משימה קשה. למשל כל ארכיטקטורה יצרה מאפיינים משלה, איך משווים בין מאפיינים שונים מי קרוב יותר? או לאיזה מדד נותנים עדיפות בציון, לחדות או לדיוק?

בעבודה המסכמת במקטע 5 אני מתאר בהרחבה כמה מדדים להשווה בין תוצרי הרשתות ומה היתרון של בעבודה זו התמקדתי ב 3 מדדים אשר משווים בין הארכיטקטרות בשתי מובנים, מדדי PSNR ו SSIM אשר משווים כמה תמונה אחת קרובה לשניה מבחינת מרחק פיקסלים. ומדד FID אשר משווה כמה מאפיינים בין מקבצי תמונות דומים.

המדדים מוסברים בהרחבה בעבודה המסכמת. אסביר בקצרה על המדדים אשר השתמשתי בעבודה זו:

Peak signal-to-noise ratio PSNR – מודד את היחס המרבי בין האות לבין הרעש, הוא נמדד בדציבלים. בהינתן תמונה I ותוצר של תמונה משופרת K ניתן למדוד את הטעות כהפרש הערך הריבועי של כל פיקסל בתמונה. החיסרון במדד זה הוא בכך שהוא לא מעניש עבור שינויים המושפעייים שונה עבור התפישה האנושית, למשל עבור דוגמא בה התמונה הוזזה בפיקסל אחד נקבל ערכים נמוכים, אפילו שבעיין האנושית לא נוכל להבחין בהבדל, זאת מכיוון שכל פיקסל עכשיו יושווה לפיקסל שלא מתאים לו מבחינת מיקום. בעבודה עשיתי שימוש במימוש של ספריית TF למדד זה.

Structural similarity SSIM - [ [5הוא מדד מרחק של ערכים סטטיסטיים הבוחן אזורים בתמונה. המדד מנסה להעריך מרחק בין שתי תמונות לפי התפיסה של מבנה התמונה כפי שמערכת הראיה האנושית מסתכלת על תמונה. המדד פועל על ידי חלוקה לאזורים והשוואה של מבני האובייקטים בתמונה התאורה, והניגודיות. המדד סימטרי ותוצאותיו המדד הם בין -1 ל 1, כאשר ככל שהתמונות דומות יתקבלו ערכים קרובים ל-1. בעבודה עשיתי שימוש במימוש של ספריית TF למדד זה.

Fréchet Inception Distance FID - [6] במדד זה משווים את התפלגויות המאפיינים של האובייקטים הנוצרים בתמונת מהרשת הגנרטיבית להתפלגויות של המאפיינים של האובייקטים בתמונות מהעולם האמיתי. ככל שהתפלגויות האיברים המיוצרים על ידי הרשת הגנרטיבית שונות מהתפלגויות האיברים בעולם האמיתי, נקבל כי המרחק יגדל. לחישוב המאפיינים של האובייקטים בתמונות משתמשים ברשת קלסיפיקציה נוספת אשר אומנה על משימת סיווג תמונות לפני השימוש בה בחישוב במדד. מעבירים מקבצי תמונות מהעולם האמיתי ומקבצי תמונות מתוצרי הרשת הגנרטיבית ברשת הנוספת, ומשתמשים בתוצאות ערכי השכבה הפנימית לחישוב ווקטורי המאפיינים של האוביקטים בתמונה. על ווקטורי המאפיינים הנוצרים מבצעים חישוב סטטיטי של מרחק בין התפלגויות (מדמים את ההתלפגויות לנורמליות לנוחות החישוב אשר משקלל את המרחק של ערכי הממוצע וערכי השונות המשותפת:

בפרויקט זה פיתחתי את מדד FID אשר משתמש ברשת inception מאומנת אשר הורד ממנה השכבה העליונה, לצורך הוצאת ווקטורי המאפיינים של התמונות.

לצורך מדידה של התקדמות האריכטקטורות או השוואה בינהם כתבתי את הפונקציה -

calculate\_fid\_ssim\_psnr\_cur\_step .

על מנת לחסוך זמן חישוב, כתבתי את הפונקציה כך שהיא קוראת פעם אחת לגנרטור לייצור התמונת מרשת הגנרטור ומחשבת את מגוון המדדים עבור אותה תמונה מיוצרת.

הפונקציה מקבלת אובייקטים מסוג tf.data.Dataset אשר נוצר בשלב ייצור הדאטה, האובייקטים מכיל מקבצים של זוגות של תמונות אשר לא הוצגו לארכיטקטורות בשלב האימון ונשמרו לצורך מדידה של טיב הארכיטקטורות והצגת התוצרים.

## SRRESNET

חיבור residual מתואר במאמר [4], הוא חיבור בין כניסת שכבה (או בלוק שכבות) למוצא שלה.

מבנה רשת הגנרטור במאמר [2] בנויה בצורת חיבורי blocks residual ו skip connections.

ארכיטקטורה במבנה שכזה מתמודדת טוב בהעלמות הגרדינאט באלגוריתם gradient descent. באלגוריתם gradient descent מנסים לחשב את כיוון שינוי המשקולות ברשת. שינוי המשקולות יהיה בכיוון השיפוע שיקטין את ערך פונקציית המטרה. חישוב הגרדיאנט בשכבות הראשונות מושפע מגרדינאטים בשכבות הסופיות ולכן ככל שמוסיפים יותר שכבות, כך השינויים בשכבות הסופיות עלול להיות קטן או חסר משמעות לשכבות הראשונות. בחיבור residual ו skip connections מדלגים על חיבור השכבות המרובות, ולכן השינוי בערכי שכבות סופיות משפיע יותר על שכבות ראשונות, וחישוב הגרדיאנט של השכבות הראשונות ביחס לסופיות מושפע בהתאם.

בנוסף, חיבור residual יוצר הקלה לרשת בייצור פונקציית הזהות במעבר בין השכבות. פונקציית הזהות היא פונקציה אשר אמור להיות קל לייצר אותה. עקרון זה מאפשר לשמור ערכים בסיסים לאורך הרשת, ורק להתאמן על ייצור שינויים תבניתיים.

בשל העקרונות הנ"ל ארכיטקטורת resnet מאפשרת להשתמש ביותר שכבות, שימוש ביותר שכבות מאפשר ליצור פונקציות מיפוי יותר מורכבות. בעבודה המסכמת אני מתאר בהרחבה ניסוי שמחברי מאמר [2] מבצעים והוא לבחון את השפעת כמות הלוקים בחיבור residual, תוצאות הניסוי הוא שככל שמגדילים את כמות הבלוקים מקבלים אימון יותר יציב ותוצאות בעלות מדדי איכות טובים יותר, ומצד שני ככל שהרשת גדלה בעקבות הוספת הבלוקים האימון נעשה איטי יותר.

בפרויקט זה מימשתי את מבנה רשת הגנרטור כפי שהיא מופיע במאמר 2, באיור 1 ניתן לראות את החיבור בין ה blocks residual ו skip connections כאשר כפי שמוצע במאמר השתמשתי ב 16 blocks residual. בכל בלוק יש חזרה של השכבות בסדר הבא: שכבת קונבולוציה , שכבת באטץ ושכבת prelu כאשר k מתאר את גודל הקרנל, n מתאר את כמות הפילטרים ו s מתאר את גודל הסטרייד.

בשלב הגדלת התמונה (הריבועים ברקע ירוק) משתמשים בשכבת pixel shuffler כפי שהוצע במאמר [11], שכבה זו מבצעת הגדלת תמונה בעזרת פרמטרים נלמדים. השכבה ממירה את האלמנטים בטנזור ממרחב העומק (כמות ה channel) למרחב גודל התמונה (גובה ורוחב). כאשר השתמשתי בשכבה זו נוצרו לי ארטיפקטים בתמונות הסופיות בצורת ריבועים. זוהי בעיה מוקרת אשר הפתרון לה הוא שינוי אופן הגדלת התמונה. בפרויקט החלפתי את בלוקי הגדלת התמונה בבלוקים שראשית מבצעים הגדלת תמונה בפרמטרים לא נלמדים על ידי שיטת nearest neighbor, אשר אמורה ליצור תמונה מוגדלת ומטושטשת, לאחר שכבת ההגדלה מבצעים שכבת קונבלוציה שאמורה ללמוד פרמטרים לאיך לשנות ולחדד את התמונה המוגדלת, רעיון זה מתואר במאמר [11] גם כן.

A diagram of a block diagram

Description automatically generated with low confidence

איור 1 : מבנה ארכיטקטורה srResNet , מורכב מבלוקים המחוברים בעקרון הresidual ו skip connections . האיור לקוח ממאמר [2]

פונקציית ההפסד בשלב זה באימון היא מרחק MSE בין פיקסלים של התמונה המקורית לתמונה הנוצרת. מרחק זה עוזר לקרב את תמונת המוצא לתמונת המקור ברמת הפיקסלים. במרחק זה מודדים את ההפרש עבור כל פיקסל בנפרד בין תמונת המטרה לתמונת הפלט מהרשת הגנרטיבית. מרחק זה מייצג את המשימה האמיתית עליה אנו נדרשים לענות, והיא לייצר תמונה כמה שיותר דומה לתמונת המקור. בפועל המרחק מודד תכונה חזקה מידי, למשל אם נחשוב על שחזור תמונת מוצא מושלם שעבר תזוזה של פיקסל אחד בודד, במקרה שכזה המרחק הנמדד יכול להיות גדול, למרות שהתמונה תראה מאוד מדויקת ואמיתית למתבונן האנושי. שימוש במרחק שכזה לרוב יוצר תמונת מטושטשת מכוון שמטרתו להגיע למינימום שגיאה של האפשרויות אשר יכול להיות הממוצע של אותן אפשרויות, ולא מחייב ערך אמיתי אפשרי.

- תמונות ברזולוציה גבוהה

- תמונות ברזולוציה נמוכה

*- הפעלת הרשת הגנרטיבית על תמונה ברזולוציה נמוכה*

באיור 2 ניתן לראות את ערכי פונקציית ההפסד של אימון srResNet , הערכים מייצגים את מרחק MSE בין הפיקסלים של תמונת המקור לפיקסלים של התמונה הנוצרת מהגנרטור. ניתן לראות כי פונקציית ההפסד לא יורדת באזור של מעל מיליון צעדים, לכן זוהי נקודה טובה לעצור. זוהי הנקודה שבה מחברי מאמר [2] עצרו את אימון SrResNet.  
A picture containing line, plot, font, text

Description automatically generated

איור 2 : גרף המציגים את פונקציית הפסד ארכיטקטורת SrResNet לאורך התקדמות האימון .

בטבלה למטה ניתן לראות את ערכי המדדים לאיכות תוצרי הרשת. מדדי PSNR ו SSIM גודלים ככל שמתקדמים באימון, עליה בערכים אלה משקפת שהקרבה לתמונת המקור גדלה מבחינת מבנה התמונה ומרחק פיקסלים.

ניתן לראות שערכי FID קטנים, ככל שערכים אלה קטנים, מתקבל כי הסטטיטסיקה של ערכי הפיצרים של האוביקטים הנוצרים דומה לסטטיטקיה של ערכי הפיצרים של אובייקטים הנוצרים מתמונות המקור. נראה כי הגרפים המוצגים מראים האטה במגמות שלהם לקראת סוף האימון ולכן אנו מצפים להאטה במגמת שיפור התוצרים של הארכיטקטורה .

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

גרפים המציגים את מדדי איכות תוצרי רשת הגנרטור כפונקציה של התקדמות האימון עבור ארכיטקטורת SrResNet . משמאל לימין, PSNR, FID, SSIM.

להלן דגימה והצגה של מקבץ תמונות אשר עברו את תהליך ההגדלה, תמונות אלו לא עברו ברשת בשלב האימון, הם נשמרו בצד לצורך הצגה. ניתן לראות בצד שמאל את התמונה המוקטנת, באמצע את תמונת המקור, ובצד ימין את התמונה המשוחזרת. כצפוי התמונות הנוצרות קרובות בבסיסן לתמונת המקור אך מטושטשטות. כפי שתואר שימוש במרחק MSE לרוב יוצר תמונת מטושטשת מכוון שמטרתו להגיע למינימום שגיאה של האפשרויות אשר יכול להיות הממוצע של אותן אפשרויות.

A blurry image of a person

Description automatically generated with medium confidenceA person with dark hair

Description automatically generated with low confidence

A person with long hair and beard

Description automatically generated with low confidenceA person with long hair and beard looking to the side

Description automatically generated with low confidence

A person with a beard

Description automatically generated with low confidenceA picture containing human face, person, person, clothing

Description automatically generated

A person smiling at the camera

Description automatically generated with low confidenceA person smiling at the camera

Description automatically generated with low confidence

A close-up of a person

Description automatically generatedA person with blonde hair

Description automatically generated with low confidence

## SrGan

באימון SRGAN משתמשים בTransfer learning לאתחול משקולות הגנרטור, תחילה מאמנים את הגנרטור ללא הדסקרימנטור בצורת SRResNet לאחר מכן ממשיכים את האימון בצורת הרשתות היריבות SRGAN. גם אני כמו במאמר [2] התחלתי את אימון SRGAN עם משקולות שנוצרו לאחר אימון SRResNet לאחר מיליון איטרציות אימון. כפי שהמאמר מציע אימנתי את ארכיטקטורת SrGan באימון של רשתות יריבות עוד 200 אלף צעדים.

בשלב זה במאמר משתמשים בשתי פונקציות הפסד אשר לחיבור ביניהם הם קוראים perceptual loss (מלשון תפיסה).

החלק הראשון הוא שימוש בcontent loss אשר הוצע במאמר [ [3ומוסבר בהמשך. מטרת חלק זה היא לשמור על מבנה התמונה, כלומר לשמור על המאפיינים הבסיסיים מתמונת המקור.

החלק השני בפונקציית ההפסד הוא adversarial lossבחלק זה נעשה שימוש בערכי תוצאות ההסתברות של הרשת הדסקרימנטיבית על תמונה שיוצרה מהרשת הגנרטיבית. מטרת חלק זה היא לגרום לארכיטקטורה לייצר טקסטורות אשר דומות לטקסטורות שנדגמו מתמונות אמיתיות אשר איתם אימנו את הארכיטקטורה.

פונקציית ההפסד של אימון הרשתות היריבות על זוגות של תמונות ברזולוציה גבוהה ורזולוציה נמוכה באימון הדסקרימטור

פונקציית ההפסד הגנרטור באימון ה adversarialהיא :

פונקציית ההפסד הכללית באימון הגנרטור :

בהצעת המאמר השתמשתי בערך

A picture containing calendar

Description automatically generated

איור 3 : הדגמת איכות התמונה ועדיפות שימוש בSRGAN אשר בעל אותה ארכיטקטורה כמו SRResNet אבל אומן בעזרת אימון רשתות יריבות. התמונה נלקחה ממאמר [2].

להלן הדגמה של תוצרי הרשת מהאימון שלי לארכיטקטורת SRGAN .

בתמונת למטה ניתן לראות בעמודות משמאל לימין את התמנות הבאות:

תמונה ברזולוציה נמוכה, התמונה המקורית ברזולוציה גבוהה, התמונה הנוצרה מארכיטקטורת SrResNet,התמונה הנוצא מארכיטקטורת SrGan. ניתן לראות כי בתמונות הנוצרו מארכיטקטורת ה SRGAN חדות בהשוואה לתמונות הנוצרו מארכיטקטורת ה SrResNet.

A blurry image of a person

Description automatically generated with medium confidenceA person with dark hair

Description automatically generated with low confidence

A person with long hair and beard

Description automatically generated with medium confidenceA person with long hair and beard

Description automatically generated with low confidenceA person with long hair and beard looking to the side

Description automatically generated with low confidence

A person with a goatee

Description automatically generated with low confidenceA person with a beard

Description automatically generated with low confidenceA picture containing human face, person, person, clothing

Description automatically generated

A person smiling at the camera

Description automatically generated with low confidenceA person smiling at the camera

Description automatically generated with low confidenceA person smiling at the camera

Description automatically generated with low confidence

A close-up of a person's face

Description automatically generatedA close-up of a person

Description automatically generatedA person with blonde hair

Description automatically generated with low confidence

איור 4 : הדגמת תוצרי הארכיטקטורות משמאל לימין התמונה ברזלוציה הנמוכה, התמונה ברזלוציה הגבוהה , ארכיטקטורת SRResNet , ארכיטקטורת SrGan. ניתן לראות כי תוצרי SrGan חדים יותר מ SrResnet. תמונות אלו נדגמו מסט הוולידציה לאחר השלמת צעדי האימון.

כאשר מתבוננים בערכי פונקציית ההפסד, ניתן לראות שיש ירידה משמעותית יותר בפונקציית הדסקרימנטור עד לשלב 1.1 מיליון צעדים. נשים לב כי מכיוןן שהדסקרימנטור הוא מסווג, אשר דרך המשוב שלו, הגנרטור לומד האם התמונות הנוצרות נראות אמיתיות או לא, רצוי כי הדסקרימנטור לא יאומן עד לאופטימליות שלו. לכן ניתן להבין למה מחברי המאמר אימנו רק עוד 200 אלף צעדים, ולא יותר. בהמשך העבודה אני מראה אימון ארוך יותר על ארכיטקטורה זו ואת תוצאותיו, בנוסף אסביר את הסיבה למה אימון דסקרימנטור באופן עדיף על הגנרטור מהווה בעיה.

A screenshot of a graph

Description automatically generated with medium confidence

איור 5: ערכי פונקציית ההפסד של הדסקרימנטור באימון SrGan. אימון בין הצעד המליון לצעד ה1.2 מליון

## בעיית קריסת מצב

קריסת מצב (mode collapse) זהו מקרה בו הארכיטקטורה מייצרת מגוון תמונות מצומצם וממשיכה לאמן ולשפר רק את המגוון התמונות המצומצם הזה.

הסיבה להגעה לקריסת מצב נעוצה באופן עדכון המשקולות של הגנרטור לפי פונקציית ההפסד לארכיטקטורת GAN, עדכון המשקולות נעשה כך שיגרום לרשת הדיסקרימנטור להחזיר הסתברות גבוהה על תוצרי הרשת הגנרטיבית, במקרה שיש אובייקטים עבורם יקבל ערכים גבוהים מאחרים, לרשת הגנרטור יהיה תמריץ להמשיך לייצר אובייקטים אלה. להרחבה בנושא, ניתן לקרוא את העבודה המסכמת שלי בסעיף 2.3.3

בשלב זה בעבודה, ניסיתי לבחון האם אכן יש צורך בtransfer learning באיתחול המשוקולות. התחלתי אימון ללא שימוש בשלב המקדים באימון בצורת SrResnetוהגעתי למקרה של קריסת מצב. הפסקתי את אימון זה לאחר חצי מיליון צעדים מכיוון שראיתי שהוא לא מתקדם בצורה טובה.

באיור 5 אני מציג בעמודה הימנית את תוצרי הרשת הגנרטיבית עבור מספר תמונות מקור המוצגות בעמודה השמאלית. ניתן לראות איך הרשת מייצרת את אותה תמונת בסיס עובר כל תמונת מקור. ניתן לראות שגוון הרקע נשמר באופן יחסי יותר נאמן לתמונת המקור, ואילו אזור הפנים בתמונה מיוצר באופן דומה יותר בין כל התמונות.

A blurry image of a person's face

Description automatically generatedA person with dark hair

Description automatically generated with low confidence

A blurry image of a person

Description automatically generatedA person with long hair and beard looking to the side

Description automatically generated with low confidence

A blurry image of a person

Description automatically generatedA picture containing human face, person, person, clothing

Description automatically generated

A blurry image of a person

Description automatically generatedA person smiling at the camera

Description automatically generated with low confidence

A blurry image of a person

Description automatically generatedA person with blonde hair

Description automatically generated with low confidence

איור 6: הצגת מקרה של קריסת מצב לאחר אימון SrGan ללא שימוש בtransfer learning

כאשר מסתכלים על ערכי מדד FID ערכו ממש גבוה, ערכו מציג את העבודה שסטטיסטיקת הפיצרים של תמונות מהעולם האמתי לאלה המיוצרות על ידי הגנרטור שונה מאוד. בנוסף ערכי ה SSIM נמוכים דבר המראה כי מבנה התמונה שונה מאוד בין תמונת המקור לתמונה המיוצרת. ערכי ה PSNR באופן יחסי גם נמוכים דבר המראה כי מרחק הפיקסלים בין תמונת המקור לנוצרת שונה.

A screenshot of a graph

Description automatically generated with medium confidence

איור 7: הצגת ערכי מדדים להשוואה בין ארכיטקטורות לאחר מקרה של קריסת מצב באימון SrGan ללא שימוש בtransfer learning .

ניתן לראות שערכי פונקציית ההפסד של הדסקרימנטור מתאפסים בצורה מהירה, במקרה שכזה לא ניתן משוב יעיל על ידי הדסקרימנטור לרשת הגנרטור. הרשת הדסקרימנטיבית למדה לפסול תמונות מהסוג הנוצר בקריסת מצב, לכן ערכי פונקציית ההפסד שלה מתאפסים. במקרה זה רשת הגנרטור לא יכולה להתקדם באימון מכיוון שהמשוב של הדסקרימנטור עבור כל תמונה הוא אותו ערך. ניתן לראות שערכי פונקציית הפסד של רשת הגנרטור עולים ככל שמתקדמים באימון.

A screenshot of a graph

Description automatically generated with low confidence

איור 8: הצגת ערכי פונקציות ההפסד של הגנרטור והדסקרימנורבשלב האיומן SrGan ללא שימוש בtransfer learning .

להדגמת המקרה של הקריסת מצב אני מייצג גרף המציג את מרחב ווקטור המאפיינים אשר מחושב ברשת הinception.

על מנת לגרום לגרף להכנס למרחב דו מימדי אני מבצע אלגוריתם PCA לקבלת וקטוריים חדשים המייצגים את מקבץ המאפיינים עם השונות המקסימילת, כאשר אני לוקח את שתי הווקטורים המציגים את השונות המקסימילית.

לצורך הדגמה של כמות הריכוזיות בקריסת מצב, אני מצג את הגרפים הבאים לפי סדר משמאל לימין, גרף התפלגויות הווקטורים לתמונת מקור, גרף בהתפלגוריות הווקטורים ממקרה קריסת המצב וגרף מהתפלגויות הווקטורים ממקרה של ארכיטקטורת SrPix2Pix שאיתה אימנתי ואתאר בהמשך.

נשים לב שהסקאלה של הצפיפות לא מנורמלת בין הגרפים, והסקאלה של מקרה קריסת המצב הוא בפקטור 10 יותר גדול, כלומר המאפיינים הרבה יותר מרוכזים באותה נקודה.

A picture containing screenshot, diagram, text, rectangle

Description automatically generatedA picture containing screenshot, colorfulness, graphics

Description automatically generatedA picture containing screenshot, colorfulness, graphics

Description automatically generated

איור 9: הצגת צפיפות ווקטורי המאפיינים של רשת inception עבור מגוון ארכיטקטורות לאחר שעברו ה המרה הורדת מימדים באלגוריתם PCA

## WGAN

מאמר [[6 בא לענות על בעייות הנובעות מהצורך לאזן בין אימוני הגנרטור והדסקרימנטור. לצורך אימון יעיל יותר של רשת הדיסקרימנטור המאמר מציע להחליף את פונקציית ההפסד של הרשת לפונקציית loss Wasserstein, מטרת פונקציית ההפסד המוצעת היא למצוא את הפונקציה אשר תביא במינימום שינוי את התפלגות של מוצא הרשת הגנרטבית להתפלגות האוביקטים מהעולם האמיתי . מרחק הווסרשטיין מספק מדד לכמות ה'"מסה" שיש להזיז, ועד כמה רחוק צריך להזיז אותה, כדי להמיר את התפלגות אחת לתפלגות השנייה. הדסקרימנטור ב-WGAN מתאמן להעריך את מרחק הווסרשטיין.

בהגדרה המקורית של אריכטקטורת GAN, הדסקרימנטור הוא מסווג, האימון של הדיסקרינטור לאופטימליות יכול לגרום לבעיה של העלמות גרדיאנטים, משום שהדיסקרימנטור עשוי להיות טוב מדי. דבר זה הופך את הלמידה של הגנרטור לקשה מכיון שלא יהיה משוב מהגרדינאטים. השינוי בפונקציית ההפסד למרחק ווסרשטיין נחוץ מכיוון שהוא מקל על הצורך באיזון האימון בין הדסקרימנטור לגנרטור. בפונקציית הפסד החדשה, ככל שהערכה של הדסקרימנטור של מרחק הווסרשטיין טובה יותר, היא יכולה להדריך את הגנרטור בלמידה לייצר דוגמאות שדומות יותר לנתונים האמיתיים. במילים אחרות, הגרדיאנט שבו הגנרטור משתמש לעדכן את הפרמטרים שלו הופך להיות מהימן יותר. מכיוון שכך ניתן לאמן את הדסקרימנטור בצורה עדיפה. מחברי המאמר מציעים לבצע 5 עדכוני דיסקרימנטור עבור כל אימון גנרטור, גם אני במעבר לארכיטקטורה זו בחרתי לבצע את אותו יחס בין עדכוני גנרטור לדיסקרימנטור.

בחישוב המרחק, על מנת למצוא את הפונקציה המינימלית שתביא למעבר בין ההתפלגויות ניתן להשתמש בדואליות Kantorovich-Rubinstein אשר מגדירה כי לחישוב המטריקה, ניתן למצוא פונקציה f אשר תקבל ערך מקסימלי עבור המרחקים

מפונקציה f ישנה דרישה להיות 1-Lipschitz רציפה ומשמע שהנורמה של הגרדיאנט שלה הוא 1 בכל מקום. הסבר מפורט יותר על מרחק Wasserstein ניתן לראות בעבודה המסכמת שלי במקטע 3.2.

מאמר [6] דואג לכך שפונקציה f תהיה 1-Lipschitz על ידי שינוי בעדכון משקולות בכל איטרציה, על ידי ביצוע תחימה של הערכים להיות בטווח מצומצם. פעולה זו גורמת לכך שהמשקולות יהיו מוכלות רק במרחב קטן ובכך ניתן לשמור את f להיות  1-Lipschitz . המאמר מציין ששיטה זו של תחימת המשקולות בעייתית מכיוון שקשה להגדיר מה התחום הנכון כפי שנסביר בהמשך בעבודה זו.

הקוד עבור ארכיטקטורה זו כתוב במחלקה SrWgan ,השינוי אשר בוצע ביחס למחלקת SrGan הוא בשינוי פונקציית האימון לדסקרימנטור ולגנרטור אשר ייגרמו לדסקרימנטור לחשב את מרחק ווסרשטיין, בנוסף לאחר כל צעד לשמירה תנאי ליפשיטץ אני מבצע את התחימה למשקולות עבור כל השכבות, השתמשתי בערכי תחימת משקולות כפי שהוצע במאמר [6] הערכים הם [0.01,0.01-[.

בגרפים למטה מוצגים ההבדלים בין תוצרי הארכיטקטורות כאשר ארכיטקטורת SrGan המקורית מוצגת בשחור והארכיטקטורה אשר עובדת עם פונקציית הפסד ווסרשטיין SrWgan, מוצגת בתכלת. ניתן לראות כי בכל המדדים להשוואה הארכיטקטורה המשתמשת בווסרשטיין עדיפה, בנוסף ניתן לראות כי האימון יציב ביחס לארכיטקטורת SrGan המקורית מכיוון שהמדדים משתנים בצורה פחותה, בייחוד FID ו SSIM אשר כמעט ולא משתנים.

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

איור 10: גרפים המציגים את מדדי איכות תוצרי הרשת כפונקציה של התקדמות האימון עבור ארכיטקטורת SrWgan בתכלת וארכיטקטורת SrGan בשחור. משמאל לימין, PSNR, FID, SSIM.

בטבלה למטה מוצגים מספר תמונת אשר נלקחו מדגימת תוצרי הארכיטקטורה תחילה מזמן 1.4 מליון צעדים ובהמשך בקפיצות של 100 אלף צעדים. התמונות מוצגות משמאל לימין, ככל שמספר הצעדים גדול כך התמונה ימנית יותר. כל זוג שורות מציגות הדגמה באימון עבור תמונה מסויימת, בשורה העליונה ניתן לראות תוצרי רשת הגנרטור בארכיטקטורת SrGan, בשורה התחתונה ניתן לראות את תוצרי רשת הגנרטור עבור ארכיטקטורת SrWGAN . ניתן לראות כי אכן המדדים לאיכות תמונה מייצגים את השינוי בתוצרי הארכיטקטורת. ככל שמתקדמים באימון SrGan בצורותו הרגילה, התוצרים מתרחקים מתמונות המקור, ונוצרים פגמים בתמונה. לעומת זאת בתוצרים של האריכטוקטורה SrWGAN, נראה כי חל שיפור ביציבות האימון, ונוצרים פחות פגמים לאורך זמן.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1.8 | 1.7 | 1.6 | 1.5 | 1.4 | Arch/steps |
|  |  |  |  |  | SrGan |
|  |  |  |  |  | SrWGAN |
| A picture containing human face, person, lip, clothing  Description automatically generated | A picture containing human face, person, lip, eyebrow  Description automatically generated | A person with long brown hair  Description automatically generated with low confidence | A person with brown hair  Description automatically generated with low confidence | A close up of a person's face  Description automatically generated | SrGan |
| A person with brown hair  Description automatically generated with low confidence | A person with brown hair  Description automatically generated with low confidence | A person with brown hair  Description automatically generated with low confidence | A person with brown hair  Description automatically generated with low confidence | A person with long brown hair  Description automatically generated with low confidence | SrWGAN |
| A close-up of a person  Description automatically generated | A person with blonde hair  Description automatically generated with low confidence | A close-up of a person  Description automatically generated | A person with blonde hair  Description automatically generated with low confidence |  | SrGan |
| A close-up of a person's face  Description automatically generated | A close-up of a person  Description automatically generated | A close-up of a person  Description automatically generated | A close-up of a person's face  Description automatically generated | A close-up of a person  Description automatically generated | SrWGAN |

איור 11: הדגמת תוצרי ארכיטקטורות SrGan וSrWgan לאורך התקדמות האימון משמאל לימין.

## WGAN-GP

כותבי מאמר [7] Improved Training of Wasserstein GANs הקלו על האימון של WGAN בכך שהורידו את הצורך בקביעת ערך לתחימה של המשקולות בכל שלב באימון כפי שנעשה ב WGAN. השיפור עוזר מפני שגודל התחימה של המשקולות הוא דבר קשה להגדרה, כפי שהסבירו במאמר [6]. תחימה קטנה מידי מקשה על אימון מהיר ותחימה גדולה מידי מקשה על שמירה של תנאי ליפשיץ הנדרש באימון. השוני בWGAN-GP הוא בהוספת תנאי רגולציה לפונקציית ההפסד, אשר עוזר לשמור על תנאי ליפשיץ. בתנאי הרגולציה החדש מודדים את הגרדיאנט של הדסקרימנטור ביחס לתמונה הנוצרת מאינטרפולציה של שתי תמונות, האחת תמונה הנדגמת מהעולם האמיתי והאחרת תמונה הנוצרת מהרשת הגנרטיבית. ככל שהגרדיאנט גדול יותר, ככה משקל התוספת של תנאי הרגולציה לפונקציית ההפסד יהיה גדול יותר.

יצרתי מחלקה חדשה אשר ממשת את אימון WGAN GP המחלקה מצויה בקובץ sr\_wgan\_gp.py

מימוש תנאי הרגולציה מבוצע בפונקציה gradient\_penatly . המימוש מבצע ראשית יצירת באטצ של אינטרפולציה בין תמונות מקור לתמונות אשר נוצרה על ידי הגנרטור, לאחר מכן מחשבים את נורמת הגרדיאנט של הדסקרימנטור ביחס לבאטץ הזה ומחשבים את העונש לפי המרחק הריבועי מהערך 1, שזהו הערך הגרדיאנט המקסימלי הרצוי עבור תנאי 1 ליפשיץ.

במאמר [11] משקל תנאי הרגולציה הוא 10 פעמים ביחס לתנאי המקורי של פונקציית ההפסד במרחק ווסרשיין. מבדיקת מטריקות לאיכות תוצרי הרשת נראה כי ארכיטקטורות SrWgan עם פונקציית הפסד ווסרשטיין ותחימת משקלות עבדה בצורה יותר טובה לאורך זמן. ניסיתי כמה פרמטרים לערכי המשקל לתנאי הרגולציה הנוסף, לרוב נראה כי אימון WGAN עם תחימת משקולות היה יותר יציב. בגרפים למטה אני מציג את התקדמות האימון לארכיטקטורת WGAN ו WGAN -GP עם כמה משקלים לתנאי הרגולציה. הגרף השחור הוא אימון WGAN עם תחימת משקולות, הגרף התכלת הוא אימון WGAN-GP עם משקל 10 אשר זהו הערך הנבחר גם במאמר [11]. הצבעים עבור כל אימון מופיעים בטבלה הבאה

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

איור12 : גרפים המציגים את מדדי איכות תוצרי הרשת כפונקציה של התקדמות האימון עבור ארכיטקטורת SrWgan ו ארכיטקטורת SrWganGp עם כמה משקלים שונים עבור תנאי הרגולציה. משמאל לימין, PSNR, FID, SSIM.

מאמר [11] ממליץ לבטל את שכבת ה batch normalization לדסקרימנטור ובמקום זאת להשתמש ב layer normalization. כאשר מיישמים batch normalization , פלט מסוים הופך להיות תלוי לא רק בקלט התואם אליו אלא גם בכל הקלטים האחרים בקבוצה, מכיוון שהממוצע והשונות מחושבים על גבי כל הקבוצה. לכן, נכון להגיד שנורמליזציה בעזרת batch normalization משנה את הבעיה ממיפוי של קלט אחד לפלט אחד למיפוי מקבוצה שלמה של קלטים לקבוצה שלמה של פלטים. בתנאי רגולציה על הגרדיאנט רצינו להטיל קנס על הנורמה של הגרדיאנט של הדיסקרימנטור ביחס לקלט יחיד, ולא לכל הקבוצה. כדי לפתור את הבעיה זו, מחברי המאמר הציעו להשמיט את שכבת ה batch normalization ולהחליפו ב layer normalization.

בשלב זה ביצעתי בדיקה כמה שינוי זה משפיע על התוצרים עבור WGANGP ואכן ישנה השפעה גדולה, התוצרים השתפרו ונהיו דומים לWGAN. בדקתי בנוסף החלפת שכבת הנורמליציה עוזרת גם בגנרטור. שם התוצרים יצאו פחות טובים. להלן הגרפים המציגים את מדדי איכות התמונה הנבדקים לאורך התקדמות האימון. ניתן לראות כי הגרפים אשר משנים את תנאי הרגולציה עבור הדסקרימנטור)תכלת סגול וכתום) דומים לגרף wgan

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

איור 13: גרפים המציגים את מדדי איכות תוצרי הרשת כפונקציה של התקמדות האימון עבור ארכיטקטורת SrWgan (שחור)ו אריכטקטורת SrWganGp בשינוי שכבות הנורמלזציה עבור הדסקרימנטור עם כמה משקלים שונים עבור תנאי הרגולציה (כתום, ורוד ,ותכלת) , SrWganGp עבור שינוי שכבת הנורמלציה גם בגנרטור עם כמה משקלים שונים (סגול וירוק) . משמאל לימין, PSNR, FID, SSIM.

בטבלה למטה ניתן לראות את ערכי המטריקות למדידת איכות תמונה וזמני הריצה עבור האימון לאחר 3 מיליון צעדים. ניתן לראות SrWgan עבד בצורה מהירה יותר, זאת מכיוון שבחישוב GP אנו קוראים לרשת נוספת לקבלת גרדיאנטים בתמונה, דבר המאט את האימון בצורה משמעותית. ניתן לראות שערכי SrWganGp10 היו הכי טובים ולכן התוצרים מצופים להיות הכי קרובים לתמונות מהעולם האמיתי.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| TIME | FID | PSNR | SSIM | ארכיטקטורה |
|  | A picture containing text, screenshot, font  Description automatically generated |  | A picture containing text, screenshot, font  Description automatically generated |  |

## Pix2Pix

ארכיטקטורת Pix2Pix אשר מתוארת במאמר [8], מאפשרת לייצר מיפוי בין תמונות לפי שינוי כללי כרצוננו. מחברי המאמר מתארים מספר שימושים לארכיטקטורה זו לצורך שינויי תמונה אשר עובדים בצורה טובה ללא צורך בהתאמת הארכיטקטורה למשימה הספציפית. להסבר מפורט עבור ארכיטקטורה זו ניתן לקרוא את העבודה המסכמת בסעיף 4.3. העקרונות העיקריים שארכיטקטורה זו מציגה הם בדסקרימנטור אשר לו הם קוראים PatchGan, שימוש בתנאי רגולציה L1 על שחזור התמונה בפונקציית ההפסד, שימוש בגנרטור מצורת UNET, ושימוש בארכיטקטורת CGAN.

דסקרימנטור Patch Gan מסווג עבור כל מקטע בתמונה האם הגיע מהעולם האמיתי או מהרשת הגנרטיבית. צורה זו של הדיסקרימנטור אשר מחזיר מספר ערכים עבור אזורים שונים מאפשרת לרשת ה GAN ללמוד הרבה יותר מידע מאשר דסקרימנטור רגיל שמחזיר תשובה כללית אחת לגבי התמונה. בצורה זו ניתנת לרשת היכולת לבחון אזורים ספציפיים כמה הם אמיתיים או לא.

בארכיטקטורת CGAN כפי שמתוארת במאמר [9] מכניסים ווקטור מידע לרשת הדסקרימנטיבית והגנרטיבית, ווקטור מידע זה מאפשר לרשת הדסקרימנטיבית לפסול או לאשר תוצרי רשת גנרטור לפי המחלקה של הווקטור. בכך אנו יוצרים התניה של איזה מחלקה אנו רוצים לייצר. בארכיקטורה Pix2Pix משתמשים בעקרון זה, לרשת הגנרטיבית והדסקרימנטיבית מכניסים תמונה שהיא הווקטור תנאי ולפי תמונה זו מייצרים את התמונה הרצויה ברשת הגנרטור.

השינויים אשר ביצעתי ליצירת המחלקה SrPix2Pix ביחס למחלקה SrGan הם:

השתמשתי בדסקרימנטור PatchGan אשר על תמונה בגדול 96X96 מחלק את התמונה ל 10X10 מקטעים שעליהם הוא בוחן האם הם הגיעו מהעולם האמיתי או מרשת הגנרטור.

הוספתי את תנאי L1 לפונקציית ההפסד בין התמונה המוקורית לתמונה המיוצרת. בדקתי כמה סדרי גודל של משקל תנאי רגולציה זה לפונקציית ההפסד.

הכנסתי את ווקטור התנאי לדסקרימנטור תמונה המקורית כפי שנעשה במאמר [[10. כאשר בחנתי את השפעת שינוי זה בנפרד על ארכיטקטורה SrGan, התוצאות היו לא טובות. בנוסף בדקתי את הארכיטקטורה SrPix2Pix ללא ווקוטר התנאי לדסקרימנטור, גיליתי שהתוצאות זהות לאימון עם ווקטור התנאי לאורך כל האימון.

התוצאות מוצגות בגרפים למטה ביחס לגרף SrWgan.

A diagram of a data flow

Description automatically generated

איור 14: מבנה PatchGan בו השתמשתי בפרויקט. שתי כניסות בגדול 96X96 ויציאה אחת בגדול 10X10

ניתן לראות כי מדד FID היה נמוך ודומה לערכי srWgan כאשר השתמשי ב L1 ביחס של 0.1 לפונקציית ההפסד עבור הגנרטור בארכיטקטורות SrGan. בנוסף ערכי ה PSNR ו SSIM היו דומים לכל ריצות ארכיטקטורת SrPix2Pix וגבוהים ביחס לארכיטקטורת SrWgan. לכן ניתן לומר כי הארכיטקטורה עבדה בצורה הטובה ביותר כאשר תנאי הרגולציה L1 בגנרטור הוסף במשקל 0.1 ביחס לפונקציית ההפסד של הגנרטור ב SrGan, בנוסף הארכיטקטורה שיפרה את תוצאותSrWgan

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

איור 15: גרפים המציגים את מדדי איכות תוצרי הרשת כפונקציה של התקדמות האימון עבור ארכיטקטורת SrPix2Pix וארכיטקטורת SrWgan עם כמה משקלים שונים עבור תנאי הרגולציה. משמאל לימין, PSNR, FID, SSIM.

בטבלה למטה ניתן לראות את תוצרי הרשתות הגנרטיביות עבור 3 מיליון צעדים מצד שמאל ארכיטקטורת SrWgan בעמודה האמצעית SrPix2Pix כאשר החלק הסופי בשם הארכיטקטורה מציג את משקל תנאי הרגולציה L1 בגנרטור, בצד ימין מוצגת תמונת המקור.

|  |  |  |
| --- | --- | --- |
| SrWgan | Pix2Pix\_01 | hr |
| A close-up of a person's face  Description automatically generated | A person with brown hair  Description automatically generated with low confidence | A person with dark hair  Description automatically generated with low confidence |
| A person with long hair and beard  Description automatically generated with medium confidence | A person with long hair and beard  Description automatically generated with low confidence | A person with long hair and beard looking to the side  Description automatically generated with low confidence |
| A person with a goatee  Description automatically generated with low confidence | A person with a beard  Description automatically generated with low confidence | A picture containing human face, person, person, clothing  Description automatically generated |
| A close-up of a person's face  Description automatically generated | A close-up of a person  Description automatically generated | A person with blonde hair  Description automatically generated with low confidence |

איור 16: הדגמת תוצרי רשת הגנרטור עבור ארכיטקטורות SrWgan ו SrPix2Pix

נסיון נוסף שביצעתי הוא לבדוק האם נוצרים קריסות מצב כאשר משתמשים בארכיטקטורת SrPix2Pix במשקולות לא מאותחלות, והאם עדיף להתחיל את האימון ממשוקולות מאופסים בהתחשב שתנאי הרגולציה L1 על תוצרי הגנרטור ימנע את קריסות המצב. בהסבר פשוט, בניסיון זה השתמשתי בארכיקטורה srPix2Pix ללא משלקולות מאומנים.

בטבלה למטה ניתן לראות את תוצאות המדדים לפי התקדמות האימון על ארכיטקטורת Pix2pix במשקל 0.1 על תנאי הרגולציה L1. התוצאות די דומות לאחר 3 מליון צעדים, עם transfer learning וללא.

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

איור 17: הצגת ערכי המדדים להשוואת ארכיטקטורות למול צעדי האימון עבור ארכיטקטורות SrPix2Pix . בתכלת אימון ללא transfer learning. בסגול ושחור אימון עם transfer learning. ניתן לראות שאין הבדל משמעותי בתוצאה הסופית בין הריצות .

## SinGan

מאמר [12] מציג אימון ארכיטקטורה המשתמשת באימון רשתות יריבות אשר מתאמן על תמונה יחידה. המאמר מציג שיטה לייצור כמה שינויים בתמונה היחידה המתקבלת. המאמר מוצג בהרחבה בעבודה המסכמת בסעיף 4.6.

הארכיטקטורה משתמשת באימן progressive growing , אשר עוזר לארכיטקטורה ללמוד מאפייני טקסטורות בכמה רמות. בכל רמה נעשה אימון WGAN GP עם דסקרימנטור PatchGan. לאחר אימון רמה מסוימת הדסקרימנטור נשאר ללא שימוש והגנרטור משומש לאימון הרמה הבאה כאשר משקולותיו נשארות מוקפאות. ישנם מספר שיטות לאימון progressive growing, העקרון הבסיסי באימון שכזה הוא שמאמנים את המודל השלם בחלקים הולכים וגדלים. הסבר על progressive growing ויתרונותיו ניתן לראות בעבודה המסכמת בסעיף 3.4. במאמר [12] מודל הדסקרימנטור נשאר בגודל קבוע בכל שלב באימון. לכן גודל המקטעים שהדסקרימנטור בוחן עליהם נשאר קבוע עבור כל גודל תמונה המיוצרת ברמות השונות, הדבר גורם לכך שכל רמה מקטעי הדסקרימנטור רואים חלקים הולכים וקטנים מהתמונה. דבר זה גורם לאימון להתמקד בכל שלב ברמת דיוק פרטים שונה. ככל שגודל הפאטץ בדסקרימנטור באופן יחיס יותר קטן בתמונה הנוצרת, הוא מתמקד באובייקטים יותר קטנים ובטקסטורות יותר חדות. ככל שהפאטץ באופן יחסי גדול הוא מתמקד באוביקטים יותר גדולים וקווי מתאר כללים.

לצורך הדגמת ארכיטקטורה זו השתמשתי בpatchGan בו השתמשתי לארכיטקטורת pix2pix, את שלב הprogressive growing ביצעתי פעמיים, לקחתי את אותה מחלקה SrWgnPix2PixHd אשר גרמתי לה לייצר הגדלה של פי 2, עשיתי זאת על ידי הוצאת שכבתupsample אחת ממודל ה resnet איתו עבדתי בפרויקט. ביצעתי 2 הגדולת תמונה פי 2 כלומר אימנתי 2 מודלי גנרטור בזה אחר זה, כאשר כל אימון גנרטור מתחיל מדסקרימנטור מאותחל במשקולות רנדומיים אשר לא משומש לאימון הבא.

תוצאות הארכיטקטורה היו פחות טובות מארכיטקטורת pix2pix בכל המדדים כפי שניתן לראות בגרפים, תוצאות הארכיטקטורה הם בצבעים כתום וסגול ואילו תוצאות ארכיטקטורת pix2pix בצבע שחור.

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

איור 18: הצגת תוצאות SinGan (גרפיים צבונעיים) למול pix2pix (בשחור), ניתן לראת כי אין שיפור בביצועים בארכיטקטורה זו למול pix2pix

## מציאת התמונות עם השיפור הגדול ביותר בין הארכיטקטורות.

לצורך הצגת ייתרון שימוש במודל נבחר, יצרתי פונקציה אשר מקבלת מספר מודלים ומבצעת חיפוש עבור התמונות שיציגו באופן הבולט ביותר את הייתרון בשימוש במודל הנבחר. הפונקציית מבצעת חיפוש בעזרת המודלים המתקבלים על תמונות ממקבץ הvalidatio. הפונקציה מקבלת כפרטמטר באיזה מדד השוואה רוצים להשתמש (המדדים הנתמכים הם FID,SSIM ). הפונקציה מחפשת את התמונות בהם ההפרש על המדד הנבחן עבור שני המודלים הראשוניים היה הגדול ביותר. לאחר מציאת התמונות בעלות ההפרש הגדול ביותר, הפונקציה שומורת את התמונות הנוצרות על ידי המודלים השונים, מספר התמונת שישמר נקבע מפרמטר. לפונקציה קוראים get\_n\_images\_with\_most\_diffrancess והיא כתובה במחלקה שאחראית על הצגת המידע בפרויקט display\_handler. להלן תוצאות ריצת הפונקציה על 2 המדדים, כאשר מודל Pix2Pix היא המודל עבורו רצינו להציג את השיפור. ניתן לראות כי אכן בתמונות המוצגרות ישנו שיפור ניכר בין ארכיטקטורת Pix2Pix לשאר הארכיטקטורת.

|  |  |  |  |
| --- | --- | --- | --- |
| SrGan | Wgan | Pix2Pix | HR |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

## הצגת התפלגות ווקטור המפאיינים ברשת inception

לצורך הדגמת התפלגות המאפיינים הנוצרים בארכיטקטורות השונות, ביצעתי תהליך של הוצאת ווקטור מאפיינים. ווקטור המאפיינים נלקחת מתוצאות מעבר תמונה בשכבה פנימית של רשת Inception מאומנת.

מכיון שאנו יכולים להציג גרפים בעבודה זו רק בדו מימד, ביצעתי התמרת PCA והצגתי עבור כל מודל שאומן גרפים אשר מכילים את 2 הווקטורים המרכזיים עבור התמרה זו. התמרת PCA מייצרת קומבינציה לינארית של הווקטרים המקוריים כך שהם מוטלים למרחב ווקטורים אורטוגונליים באופן בו הווקטור המרכזי הוא בעל השונות הגדולה ביותר עבור המאפיינים המקוריים, והווקטור השני הוא בעל השונות השניה בגדולה עבור ווקטור המאפיינים.

ניתן לראות את הפילוג של תמונות אמיתיות, ופילוג של תמונות הנוצרות בארכיטקטורות השונות. כפי שתיארנו בחלק של קריסת המצב בסעיף 8 בעבודה זו, אכן ניתן לראות שבקריסת המצב כל התמונות הנוצרות היו מאותם מאפיינים דומים מכיוון שהערכים מתרכזים רק באזור מצומצם מאוד בגרף.

|  |  |  |  |
| --- | --- | --- | --- |
| Pix2Pix | SrGan | SrGan No Transfer | HR |
|  |  |  |  |

## סיכום

בפרויקט זה מימשנו ארכיטקטורות GAN המבצעות סופר רזולוציה.

הסברנו עקרונות אימון לארכיטקטורת GAN, הסברנו את האלגוריתם הבסיסי והסיבות לקושי באימון.

מטרת הפרויקט היתה לראות איך שינויים אשר הוצעו במאמרים לארכיטקטורות GAN יכולים להשפיע על תוצאות מאמר [2]. ראינו הרבה שינויים אשר משפרים את הארכיטקטורה המקורית עליה עבדנו כמודל בסיסי SrGan .

במאמר הראנו שימוש במדדים שונים אשר התאימו לתוצאות השונות שקיבלנו.

ניתן לומר כי לדסקרימנטר PatchGan ופונקציית הפסדWasserstain היתה השפעה ניכרת על התוצאות ארכיטקטורת SrGan.

בפרויקט זה היה מספר נושאים אשר עליהם לא דיברנו וניתן לחשוב עליהם כהמשך לפרוירקט.

לרוב שמייצרים מאפיינים חדשים, רוצים ליצור מגוון שלהם, בהרבה ממאמרי GAN מתיחסים למידת הגיוון במאפיינים, ניתן לבחון האם אצלנו נוצרים מאפיינים מגוונים חדשים ואיך ניתן להשפיע על מידת הגיוון.

בפרויקט דנו בהגדלת רזולוציה עד פי 4 כפי שבוצע במאמר [2], ניתן לבחון האם הגדלת רזולוציה בפקטור יותר גדול מצריכה פתרונות אחרים, או שהארכיטקטורות שהצגנו מספקות.

ביצוע שיפור ברזולוציית התמונה הוא אלגוריתם שימושי מאוד, שיפור האלגוריתמים בו יכול להיות בכמה מובנים כגון מהירות הפעלה, יציבות האימון, דיוק התוצאות וחדות התמונה. כאשר ביצענו השוואות בין ארכיטקטורות בעבודה זו, אלו המאפיינים אשר דנו בהם.

## רשימת מקורות

[1[ Goodfellow, Ian, et al. "Generative adversarial networks." Communications of the ACM 63.11 (2020): 139-144.‏

[2] LEDIG, Christian, et al. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 4681-4690.

[3] JOHNSON, Justin; ALAHI, Alexandre; FEI-FEI, Li. Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision. Springer, Cham, 2016. p. 694-711.

[4] HE, Kaiming, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770-778.‏

[5] Wang, Zhou, et al. "Image quality assessment: from error visibility to structural similarity." IEEE transactions on image processing 13.4 (2004): 600-612

[6] Arjovsky, Martin, Soumith Chintala, and Léon Bottou. "Wasserstein generative adversarial networks." International conference on machine learning. PMLR, 2017.‏

[7] Gulrajani, Ishaan, et al. "Improved training of wasserstein gans." Advances in neural information processing systems 30 (2017).‏

[8] Isola, Phillip, et al. "Image-to-image translation with conditional adversarial networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.‏

[9] Mirza, Mehdi, and Simon Osindero. "Conditional generative adversarial nets." arXiv preprint arXiv:1411.1784 (2014).‏

[10] Qiao, Jiaojiao, et al. "Image super‐resolution using conditional generative adversarial network." IET Image Processing 13.14 (2019): 2673-2679.‏

[11] Shi, Wenzhe, et al. "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.‏

[12] SHAHAM, Tamar Rott; DEKEL, Tali; MICHAELI, Tomer. Singan: Learning a generative model from a single natural image. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019. p. 4570-4580.‏