**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. 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. כך 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

מבוא

הסבר על SRGAN

הסבר על פונקציית ההפסד

הסבר על אימון רשתות GAN

הסבר על מבנה הארכיטקטורה ResNet

תוצאות SRGAN SRRESNET

שימוש ב WGAN

שימוש ב -WGANGP

שימוש ב CGAN

שימוש ב PATCH-GAN

הסבר תוצאות FID SSIM PSNR

מבנה המחלקות

# 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 של 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.

בעבודה אתבסס ראשית על מאמר [2] אשר הציע את ארכיטקטורתSRGAN , אשר 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.

האלגוריתם ממאמר [2] דורש באימון תמונות באיכות רזולוציה גבוהה ותמונות באיכות רזולוציה נמוכה, על כך המחברים מציעים לייצר זוגות של תמונות. ניתן לייצר תמונה מוקטנת מתמונה מוגדלת בכמה דרכים כמו למשל על ידי החלקה של התמונה המוגדלת ודגימה כפי שהמחברים עשו במאמר.

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. אופן נוסף לבדיקת טיב הארכיטקטורה הוא 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 מולדים 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 האיומן לגרפים השונים
* Keras הוא API של רשתות עמוקות ברמה גבוהה, 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 באמצעות 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 באופן יעיל.

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. גדלי התמונת המקוריות הן 218X178.

לטעינת התמונת ייצרתי מחלקה עוטפת הנקראת MyDataLoader הנמצאת בקובץ DataLoader. המחלקה מייצרת מקבצים (באטצ'ים) של טנסורים בעזרת שימוש בתשתית הנינתנת 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 על תמונות אלה לא יבוצע אימון, תמות אלה ישומשו לבידקת איכות תוצרי הארכיטקטורות השונות.

בעבודה זו השתמשתי בפרמטרים לאימון הדומים לפרמטרים לאימון מהמאמר [2]. גודל הבאטץ 16, גודל התמונה ברזולוציה הגבוה 96, יחס ההקטנה לרזולוציה הנמוכה 4, בנוסף השארתי 5 אחוז 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 מ2 תמונות אחת ברזלוציה גבוהה והשניה ברזלוציה נמוכה.

## הסבר על אופן מדידה והשוואה בין הארכיטקטורות.

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. לצורך שיפור איכות התמונה האלגוריתמים מוסיפים מידע וטקסטורות אשר למדו בשלב האימון מאוביקטים מהעולם האמיתי.

אין הגדרה מסויימת איך התמונות הסופיות בתוצרי הרשת אמורים להראות, לכן השוואה בין ארכיטקטורת לרשתות יריבות זו משימה קשה.

בעבודה המסכמת במקטע 5 אני מתאר בהרחבה כמה מדדים להשווה בין תוצרי הרשתות ומה הייתרון של כל אחד על פני האחר.

לצורך מדידה והשוואה בין תוצרי הרשתות שאייצר אשתמש בכמה מדדים לבדיקת טיב איכות תוצרי הרשת אשתמש במדדים הבאים אשר אסביר פה בקצרה, ומוסברים בהרחבה בעבודה המסכמת:

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 עישתי 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 ל1. בעבודה עישתי 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 המויצרים על ידי הרשת הגנרטיבית שונות מהתפלגויות האבירים 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, 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 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.

הפונקציה מקבלת אויבטקtes\_dataset מסוג tf.data.Dataset אשר נוצר בשלב ייצור הדאטה, האויבטק מכיל באטצים של זוגות של תמונות אשר לא הוצגו לאריכטקטורות בשלב האימון ונשמרו לצורך מדידה של טיב האריכטקטורות 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 XX 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 שכבת קונבולוציה , שכבת באטץ ושכבת prelu כאשר k מתאר את גודל הקרנל, n מתאר את כמות הפילטרים ו s מתאר את גודל הסטרייד.

ניתן לראות שלהגדלת הרשת מתשמים ב שכבת pixel shuffler X2 אשר יצר לי מבנים ריבועיים בתוצרי הגנרטור, אותו החלפתי לשכבת upsample שונה

A diagram of a block diagram

Description automatically generated with low confidence

איורXX : מבנה אריכטוקרה SRRESNT , מורכב מבלוקים המחבורים בעקרון הresidual ו skip connections . האיור לקוח ממאמר[2]

בפונקציית ההפסד בשלב זה באימון משתמשים במרחק MSE בין פיקסלים של התמונה המקורית לתמונה המגונרטת. 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 XX ניתן לראות את ערכי פונקציית ההפסד של אימון SRGAN

A picture containing line, plot, font, text

Description automatically generated

להראות פה הדגמה של תחילת האימון לתומונות הנוצרות סטפים קטנים

באיורים אלו ניתן לראות את ערכי המדדים PSNR ו SSIM גודלים ככל שמתקדמים באימון, עליה בערכים אלה משקפת שהקרבה לתמונת המקור גדלה מבחינת מבנה התמונה ומרחק פיקסלים.

ניתן לראות שערכי FID גדלים גם, ככל שערכים אלה גדלים מתקבל כי הסטטיטסיקה של ערכי הפיצרים של האוביקטים הנוצרים שונה מהאוביקטים של תמונות המקור. הסיבה לכך כנראה מכיון שהתמונות נהיות יותר מטושטושות עם הזמן, בעוד שתמונות המקור חדות.

A screenshot of a graph

Description automatically generated with medium confidence

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 וכפי שתואר שימוש במרחק שכזה 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

SRAGN

באימון SRGAN משתמשים בTransfer learning לאתחול משקולות הגנרטור, תחילה מאמנים את הגנרטור ללא הדסקרימנטור בצורת SRResNet לאחר מכן ממשיכים את האימון בצורת הרשתות הירביות 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 הפיצ'רים הבסיסיים בתמונת 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

באיור XX הלקוח ממאמר [2] רואים את ההבדל בין שימוש ב SRResNet ל SRGAN אשר שניהם תוצרים של אותה ארכיטקטורת גנרטור. ארכיטקטורת SRGAN אומנה בתוספת של פונקציית הפסד adversarial loss. ניתן לראות שאכן התמונה של SRGAN נוצרת חדה יותר. הסיבה לכך היא שהרשת הגנרטבית מייצרת תמונות אשר יקבלו הסתברות גבוהה ברשת הדסקרימנטיבית אשר אומנה לסווג האם תמונה הגיעה מהעולם האמיתי או מהרשת הגנרטבית. שימוש בעקרון זה מייצר תמונת אשר נראות יותר דומות לתמונות מהעולם האמיתי, והדבר לפעמים מגיע על חשבון טעות בדיוק של שחזור הפרטים כפי שניתן לראות באיור. האיור מציג בנוסף את ערכי PSNR(db) וערכי SSIM על גבי כל תמונה שהגדילו לה את הרזולוציה, כפי שתיארנו ערכים אלה מודדים מרחק בין פיקסלים לתמונת המקור, אשר נותן פחות משקל לאיכות התמונה ויותר משקל לכמה התמונה קרובה לתמונת המקור.

ניתן לראות שהטקסטורה ב SRGAN אומנם חדה ונראית אמיתית, אבל לא נאמנה למקור,למשל באזור הבד שעל המצח והצוואר.

A picture containing calendar

Description automatically generated

איור XX : 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 .

בתמונה XXX 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 בהשווה 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 XX : הדגמת תוצרי הארכיטקטורות משמאל לימין התמונה ברזלוןציה 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.

ערכי 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

## 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.

בתמונה XX 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

ערכי המדדים  
FID ממש גבוה, 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 למצוא שונה.

A screenshot of a graph

Description automatically generated with medium confidence

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 של הקריסת 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 ההפסד של רשת הגנרטור עולים ככל שתמקדמים באימון.  
A screenshot of a graph

Description automatically generated with low confidence

### 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.

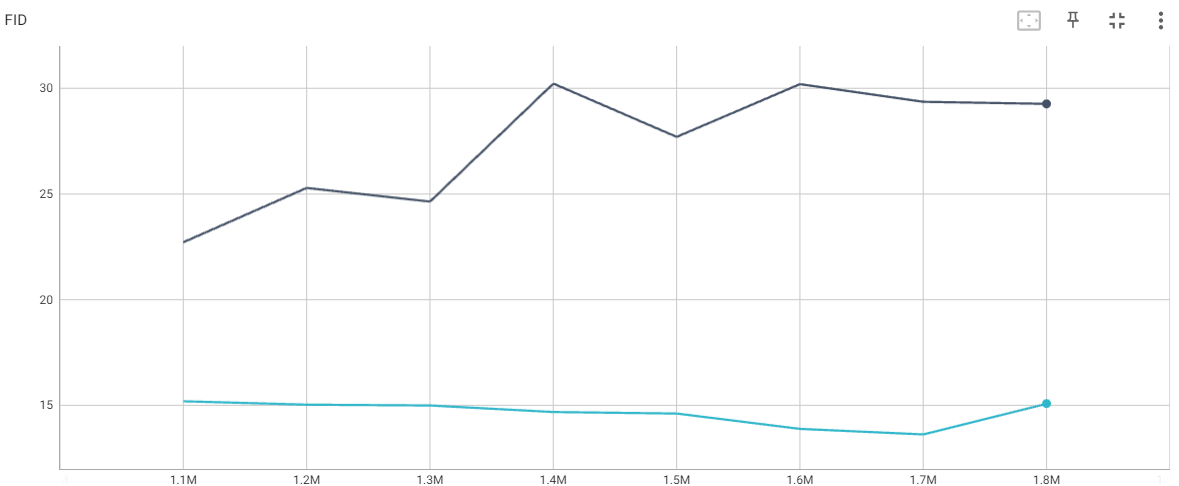
בהגדרה המקורית של אריכטקטורת GAN, הדסקרימנטור הוא מסווג, האימון של הדיסקרינטור לאופטימליות יכול לגרום לבעיה של העלמות גרדיאנטים, משום שהדיסקרימנטור עשוי להיות טוב מדי. דבר זה הופך את הלמידה של הגנרטור לקשה מכיון שלא יהיה משוב מהגרדינאטים. השיוני בפונקציית ההפסד למרחק ווסרשטיין נחוץ מכיוון שהוא מקל על 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.

קוד המחלקה נמצא בקובץ sr\_wgan\_clip ,השינוי אשר בוצע ביחס למחלקת sr\_gan הוא בשיוני 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 אני מראה את ההבדלים בין תוצרי האריכטקטורת כאשר ארכיטקטורת SrGan המקורית מוצגת בשחור והארכיטקטורה אשר עובדת עם פונקציית הפסד ווסרשטיין מוצגת בתכלת. ניתן לראות כי בכל המדדים לאיכות תמונה הארכיטקטורה המשתמשת בווסרשטיין עדיפה, בנוסף לראות כי האימון יציב ביחס לארכיטקטורת SrGan המקורית מכייון שהמדדים משתנים בצורה פחותה, בייחוד FID ו SSIM אשר כמעט ולא משתנים. A picture containing line, diagram, plot, origami

Description automatically generated A picture containing line, diagram, plot

Description automatically generated

הדגמת מדדי ארכיטקטורות SrGan לאורך התקדמות. שחור מייצג את אימון SrGan ותכלת מייצג את הארכיטקטורה בתוספת ווסרשטיין.

בטבלה למטה אני מציג מספר תמונת אשר נלקחו מדגימת תוצרי הארכיטקטורה מזמנים 1.4 מליון צעדים בקפיצות של 100 אלף צעדים. התמנות 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 תצורי רשת הגנרטור באותה ארכיטקטורה כאשר שיניתי את פונקציית ההפסד לווסרשטיין נקרא לאריכטקטורה זו 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 בשינוי פונקציית ההפסד לווסרשטיין, נראה כי חל שיפור בייציבות האימון, והפגמים פחות נוצרים.

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

הדגמת תוצרי ארכיטקטורות SrGan לאורך התקדמות האימון משמאל לימין. בשורות למעלה פונקציית ההפסד המקורית ובשורות למטה פונקציית ההפסד ווסרשטיין.

## Conditional GAN

C:\Users\liorb\Desktop\work\SrProject\save\_work\save\_sr\_gan

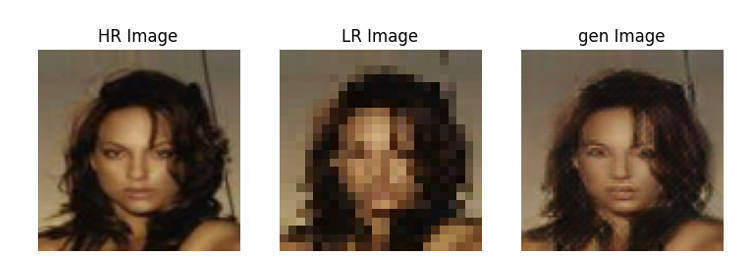
לפני החלפת UPSAMPLE

אומנם התמונה נהיית יותר חדה

אבל, ניתן לראות ריבועים בתמונה המשוחזרת

בנוסף כנראה שהאימון היה לא תקין כי ה GEN DIS LOSSES היו 0 די כל הזמן  
A picture containing human face, person, glasses, forehead

Description automatically generated



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.‏

A picture containing human face, person, person, clothing

Description automatically generatedA person with a goatee

Description automatically generated with medium confidence

A person with brown hair

Description automatically generated with low confidenceA close-up of a person's face

Description automatically generated



A close-up of a person smiling

Description automatically generatedA person smiling at the camera

Description automatically generated with low confidence

A close-up of a person smiling

Description automatically generatedA blurry image of a person smiling

Description automatically generated with medium confidence

A close-up of a person smiling

Description automatically generatedA person smiling at the camera

Description automatically generated with low confidence

A person with dark hair

Description automatically generated with low confidenceA person with brown hair

Description automatically generated with low confidence

A person with dark hair

Description automatically generated with low confidenceA blurry image of a person

Description automatically generated with medium 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 long hair and beard

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

Description automatically generated with low confidence

A person with long hair

Description automatically generated with medium confidenceA person with long hair

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

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

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

Description automatically generated with low confidenceA person with long hair

Description automatically generated with medium confidenceA person with long hair

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

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

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

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

Description automatically generated with low confidenceA person with long hair

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

Description automatically generated with low confidenceA person with blonde hair

Description automatically generated with low confidenceA person with blonde hair

Description automatically generated with low confidenceA close-up of a person wearing a baseball cap

Description automatically generated with medium confidenceA person wearing a baseball cap

Description automatically generated with low confidenceA blurry image of a person

Description automatically generated with medium confidenceA blurry image of a person

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

Description automatically generated with low confidenceA close-up of a person's face

Description automatically generatedA close-up of a person's face

Description automatically generatedA person with long hair

Description automatically generated with low confidenceA person with long 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

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

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

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

Description automatically generated with low confidenceA 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

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

Description automatically generated with low confidenceA 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

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

Description automatically generated with medium confidenceA person with long hair

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

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

Description automatically generated with low confidenceA 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

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

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

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

Description automatically generated with medium confidenceA close-up of a person's face

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

Description automatically generatedA person with brown hair

Description automatically generated with low confidenceA person with brown hair

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

Description automatically generatedA person with long brown hair

Description automatically generated with low confidenceA person with brown hair

Description automatically generated with low confidenceA close-up of a person's face

Description automatically generatedA close-up of a person's face

Description automatically generatedA close-up of a person

Description automatically generatedA person with brown hair

Description automatically generated with low confidenceA person with long brown hair

Description automatically generated with low confidenceA person with long brown hair

Description automatically generated with low confidenceA close-up of a person's face

Description automatically generatedA close-up of a person's face

Description automatically generatedA blurry image of a person

Description automatically generated with medium confidenceA blurry image of a person

Description automatically generated with medium confidenceA close-up of a person's face

Description automatically generatedA close-up of a person's face

Description automatically generated לתאר את הבעיה -- שיטות פתרון -- בכל אחד רקע קצר אם הפניה לעבודה המסכמת  
את הדרך שאני בוחר

את הכלים

למה בחרתי בדרך הזאת

מה הרצתי

ניתוח כלים שהשתמשתי בהם

מסקנות

השוואה בין שיטותשונות מסכנות שונות

מבנה הפרויקט

I

IDENTETY MAPING?  
  
CONTINAL GAN  
PATCH GAN

להסביר על SRGAN

להראות תוצאת PSNR SSIM FIDככל שמתקתמים באימון

להראות שימוש ב WGAN GP

להראות שימוש ב CGAN

להראות שימוש ב PATCH GAN

להוסיף שימוש ב L1 LOSS RECONSTURCTION

לעשות נירמול לדאטה 1-1 01

לעשות דינירמול

לעשות באטץ נומליזשן מכובה בTEST

להראות איך השפעה של UPSCALE שונה

להראות אימון ארוך WGAN אם עוזר או לא לעומט GAN רגיל שאמור להתקע ולפעמים לעלות ולרדת באימון

להראות LOSS חדש פשוט L1 ממוצע (כמו PSNR)

להראות BICUBIC INTERP

להראות לוס מול DOWNSCALE L1

https://github.com/henry32144/wgan-gp-tensorflow

https://arxiv.org/pdf/1705.02438v1.pdf#page=4&zoom=100,165,946

<https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/iet-ipr.2018.6570>

docker run --gpus all -it -v /home/ubuntu/srgan:/tf/srgan tensorflow/tensorflow:latest-gpu bash

תיקון GP

https://chat.openai.com/share/a5a4dcc6-837d-4e1d-b077-7c7287f67ba9