# Review 12: Alias-Free Generative Adversarial Networks

**Paper: https://arxiv.org/abs/2106.12423v4**

פינת הסוקר:

המלצת קריאה ממייק: חובה לעוסקים במודלים גנרטיביים של הראייה הממוחשבת, לכל האחרים מומלץ מאוד.

בהירות כתיבה: גבוהה מינוס.

רמת היכרות עם כלים מתמטיים וטכניקות של ML/DL הנדרשים להבנת מאמר: היכרות עם עקרונות של GAN-ים, הבנה של טכניקות דגימה (downsampling, upsampling) ושחזור אות רציף מדגימותיו (משפט דגימה של נייקוויסט, נוסחת שנון-וויטקר).

יישומים פרקטיים אפשריים: יצירה של תמונות equivariant להזזה ולסיבוב ממרחב לטנטי של GAN.

פרטי מאמר:

פורסם בתאריך: 23.06.21, בארקיב.

הוצג בכנס: טרם ידוע.

תחומי מאמר:

גאנים

מניעת פיקסלים ״קפואים״ (דבוקים) למקום בתמונות מגונרטות.

הקטנה של aliasing בתמונות המגונרטות באמצעות גאנים.

כלים מתמטיים, מושגים וסימונים:

StyleGAN2

Translation/rotation equivariance

התמרת פוריה (Fourier transform)

Aliasing

נוסחת אינטרפולציה של Whittaker–Shannon

מסננים לשחזור אות רציף מדגימותיו (sinc, jinc, מסנן קייזר)

מבוא:

בשנים האחרונות איכות ורזולוציה של תמונות, הנוצרות באמצעות GAN-ים השתפרו משמעותית. ארכיטקטורות גאנים שונות הצליחו ליצור תמונות באיכות מדהימה ובעלות רזולוציה גבוהה עבור מגוון משימות של הראייה הממוחשבת בדומיינים רבים כגון:

יצירה של תמונות פנים ותמונות פוטוריאליסטיות אחרות.

יצירת דמויות מצוירות (anime).

"העתקה תמונה" לדומיין אחר (כמו יצירת תמונה מסקיצה, שינוי סגנון תמונה, שינוי של דמויות בתמונה וכדומה).

יצירה תמונה מתיאור מילולי.

בנוסף הוצעו ארכיטקטורות כמו StyleGAN2 המסוגלות ליצור פיסות דאטה ויזואלי בעלות פיצ'רים ויזואליים נתונים (disentangled) כגון גיל, צבע שיער, צורה של גבות וכדומה. למרות כל ההצלחות המרשימות האלו נותרו מספר שאלות בנוגע לעקרונות של תהליך יצירת תמונות באמצעות רשתות נוירונים.

המאמר מציין כי פיצ'רים בעלי סקאלות (scales) שונות בתמונות טבעיות הן בעלי מבנה היררכי מובהק. כלומר הזזה של ראש בתמונה אמורה לגרום לשיער לזוז בצורה דומה. לכאורה נראה כי מנגנון של יצירת תמונות בגאנים אמור לבנות תמונות עם פיצ'רים בעלי מבנה היררכי דומה. למשל רשת הגנרטור של גאן (כגון StyleGAN) מתחילה מיצירת תמונה ברזולוציה נמוכה ואז מבצעת פעולת upsampling כדי ליצור תמונות ברזולוציה גבוהה יותר.

תיאור הבעיה:

המאמר הנסקר טוען כי למרות הדמיון לעיל (בין תהליכי יצירת תמונות) הפיצ'רים הגסים בתמונות, הנוצרות באמצעות גאנים,שולטים רק ב"נוכחות" (נראות) של הפיצ'רים העדינים ולא במיקום שלהם. אי קוהרנטיות זו באה לידי ביטוי כאשר מזיזים או מסובבים פרט גדול בתמונה הנוצרת באמצעות גאן (באמצעות שינוי של הייצוג הלטנטי). המחברים מראים כי במקרים רבים ניתן לראות פרטים עדינים (המהווים חלק של הפרט הגדול) של התמונה שקופאים באותו מקום בתמונה במקום לזוז/להסתובב יחד עם הפרט הגדול. המאמר מכיל מספר דוגמאות לתופעה המתוארת לעיל: כמו פרווה סביב העין של חתול נשארת במקום כאשר מזיזים את העין, השיער לא זז כאשר משנים את תנוחת הראש ודוגמאות רבות אחרות לכך. נציין, כי קיום פיקסלים קפואים/דבוקים כאלו מעיד על העדר equivariance לפעולות הזזה וסיבוב של הגנרטור.

תמצית מאמר:

המאמר מראה כי מניעת התקפלות תדרים (aliasing) מקלה את בעיית equivariance לסיבוב של תמונות הנוצרות ע״י הגנרטור. עבודות קודמות מציינות כי התקפלות תדרים בתמונות, הנוצרות באמצעות רשתות נוירונים ובפרט ע״י גאנים, היא תופעה הנגרמת מפעולות לא לינאריות ופעולות downsampling לא מדויקות כמו pooling או strided convolution. התקפלות תדרים מייצרות תדרים מעבר לתדר נייקוויסט (Nyquist)/מחצית תדר הדגימה של התמונה (הנגזר מהרזולוציה שלה). אותם תדרים שאינם מפולטרים מתקפלים לתדרים הנראים ו״מתחזים״ לתדרים אמיתיים אף שאינם קיימים במקור. כתוצאה מכך השכבות הבאות של הרשת עלולות ״ללמוד פיצ'רי שווא״ המסתמכות על הארטיפקטים הנוצרים עקב התקפלות תדרים (ראה An Effective Anti-Aliasing Approach for Residual Networks להסבר מעמיק יותר על הקשר שתואר לעיל).

המחברים מציעים להשתמש במסננים מעיבוד אותות למניעת התקפלות ומראים כי מסננים אלו מצליחים לגרום לתמונות המגונרטות באמצעות הגנרטור להיות equivariant לפעולות הזזה לסיבוב. נציין כי כדי לגרום לתמונה להיות equivariant לסיבוב המאמר משתמש במסנן שהוא רדיאלי-סימטרי (radial-symmetric) בעל תגובת תדר בצורת דיסק.

תקציר מאמר:

המאמר מציין שתי סיבות תופעת התקפלות תדרים המתוארת בפסקה הקודמת:

שימוש במסננים (פילטרים) כגון בילינארי או strided convolution במהלך יצירת תמונה.

שימוש בפונקציות אקטיבציה לא לינאריות הפועלים על כל פיצ'ר בנפרד.

המאמר הנסקר מציע לנצל שיטות anti-aliasing קלאסיות מתחום עיבוד אותות להתמודדת עם התקפלות תדרים בתמונות. למעשה המחברים מתייחסים לתמונה כאל דגימה של אות דו-מימדי רציף. אות רציף זה הוא בעל רוחב פס סופי (bandlimited) מאחר והוא צריך להיות מיוצג בצורה נאמנה באמצעות דגימה בגריד (grid) של פיקסלים. כותבי המאמר טוענים כי שימוש בטכניקות anti-aliasing במהלך יצירה של תמונה ע"י הגנרטור מצליחה להקטין את חוסר equivariance בתמונה הנוצרת באופן משמעותי. המחברים טוענים כי טכניקות אלו מאפשרות למנוע מפיצ'רים ויזואליים עדינים של התמונה הנוצרת להיות "דבוקים" למיקומים קבועים בתמונה ובכך נפתרת בעיית הפיקסלים השרופים.

כאמור המאמר מזהה שתי סיבות להתקלפות תדרים שמופיעה בתמונות שהגנרטור יוצר: שימוש במסננים לא מדויקים ופונקציות אקטיבציה לא לינאריות המופעלות ברמה של פיצ'ר, שעלולים ליצור תדרים ״גבוהים מדי״. המחברים מציעים לשנות את הארכיטקטורה של הגנרטור (הדיסקרימינטור נותר ללא שינוי) באופן הבא:

החלפת קונבולוציות 3x3 ב-StyleGAN2 המקורי בקונבולוציות 1x1 סימטריות שבאופן די ברור equivariant לסיבוב (נציין כי קונבולוציות 3x3 הן equivariant להזזה אך לא לסיבוב).

הוספה של מסנן upsampling בפקטור m (הכנסת אפסים בין הדגימות) לפני כל אקטיבציה לא לינארית (Leaky ReLU) ולאחריה מסנן downsampling באותו פקטור m בכל שכבה של הגנרטור. מכיוון שכל שכבה של הגנרטור מבצעת upsampling של תמונה (הגדלת רזולוציה פי שתיים) לפני הפעלת האקטיבציה, ניתן לאחד אותו עם ה-upsampling בפקטור m הנדרש עבור ״טיפול בפונקציית אקטיבציה״ ולבצע upsampling בפקטור 2m. מעשית המאמר משתמש ב-m=2. המסנן ה-downsampling הנבחר הוא מסנן קייזר עם המותאם כי להיות רדיאלי-סימטרי.

איך מודדים equivariance להזזה ולסיבוב:

כדי לשערך את מידת ה-equivariance של התמונה X, הנוצרת באמצעות מוקטור לטנטי w, להזזה/סיבוב, יש להבין איזו טרנספורמציה צריך לעשות ל-w, כדי להזיז/לסובב את התמונה בהיסט/זווית נתונה t. חיפוש אחרי טרנספורמציה כזו עבור ארכיטקטורה הסטנדרטית של StyleGAN2 הוא די מורכב. כדי להתמודד עם סוגייה זו, המחברים מציעים להחליף את הקלט הקבוע לרשת ה-synthesis בפיצ'רי פוריה. נזכיר כי רשת ה-synthesis בונה את התמונה משני קלטים:

וקטור הסגנון w שנבנה מוקטור לטנטי z מהתפלגות גאוסית באמצעות העברתו של z דרך רשת מיפוי (mapping network). תת-וקטורים של w ״מוזרקים״ לשכבות שונות של רשת ה-synthesis ליצירת פיצ'רים בסקאלות שונות.

וקטור דטרמיניסטי נלמד w0.

אז כדי להקל על חיפוש טרנספורמציה של w0 שתגרום הזזה/סיבוב של תמונה בהיסט/זווית נתונה t, המחברים מציעים להחליף את w0 בפיצ'רי פוריה. התדרים של פיצ'רי פוריה אלו נדגמים (פעם אחת ונותרים קבועים בהמשך) מהתפלגות אחידה מפס התדרים המתאים לתמונה ברזולוציה הנמוכה ביותר שיש ב-StyleGAN2 (כלומר 4x4). החלפה זו מאפשרת למצוא את הטרנספורמציה T ל-w0 כדי שהגנרטור ייצור מ-(T(w0 את התמונה המוזזת/המסובבת, בצורה קלה.

לבסוף השערוך של equivariance של התמונה X להזזה/סיבוב נמדד באמצעות (peak signal-to-noise ratio (PCINR בין X לתמונה המוזזת/המסובבת ״המושלמת״ לבין זו, הנוצרת מהווקטור הלטנטי המוזז/המסובב. לבסוף equivariance מחושב בתור ממוצע של PCINR-ים מעל סיבובים/הזזות של ווקטורי w האפשריים.

בנוסף המאמר מציע כמה שינוים לארכיטקטורה של StyleGAN2 שביניהם ביטול רעש פר-פיקסל והקטנה של מספר השכבות בגנרטור.

הישגי מאמר:

המאמר הצליח ליצור תמונות שהם משמעותית יותר equivariant לפעולות סיבוב ולהזזה מאלו הנוצרות באמצעות StyleGAN2 הסטנדרטי תוך שמירה על אותו FID.

נ.ב.

אחד המאמרים הראשונים שהציע שיטה מבוססת טכניקות anti-aliasing קלאסיות להתמודד עם תופעת הפיקסלים הדבוקים, המתרחשת בתמונות הנוצרות באמצעות StyleGAN2. אני מניח שמאמר זה יהווה סנונית ראשונה למחקרים בנושא של התמודדות עם תופעת התקפלות תדרים במודלים גנרטיביים בתחום הראייה הממוחשבת .

ברציוני להודות עמוקות לישי טובי, מנכל טכנולוגי ב EntityMed.com על עזרתו בפיענוח עיבוד האותות במאמר. ותודה ענקית ללירון יצחקי, ראש צוות מחקר HourOne.ai על סיעור המוחין בהבנת המאמר. ללא עזרתם של ישי ושל לירון סקירה זו לא היתה יוצאת לאור!!