

סקירה זו היא חלק מפינה קבועה בה אני סוקר מאמרים חשובים בתחום ה-ML/DL, וכותב גרסה פשוטה וברורה יותר שלהם בעברית. במידה ותרצו לקרוא את המאמרים הנוספים שסיכמתי, אתם מוזמנים לבדוק את העמוד שמרכז אותם תחת השם [deepnightlearners](#).

---

לילה טוב חברים, היום אנחנו שוב בפינתנו deepnightlearners עם סקירה של מאמר בתחום הלמידה העמוקה. היום בחרתי לסקירה את המאמר שנקרא:

## Alias-Free Generative Adversarial Networks

---

### 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:

**לינק למאמר:** [זמין להורדה](#).

**לינק לקוד:** [הגיט](#) אומר שיצא בספטמבר.

**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 לשיעור 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) הפיצורים הגסים בתמונות, הנוצרות באמצעות גאנים, שולטים רק ב"נוכחות" (נראות) של הפיצורים העדינים ולא במקום 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.



Figure 1: Examples of “texture sticking”. Left: The average of images generated from a small neighborhood around a central latent (top row). The intended result is uniformly blurry because all details should move together. However, with StyleGAN2 many details (e.g., fur) stick to the same pixel coordinates, showing unwanted sharpness. Right: From a latent space interpolation (top row), we extract a short vertical segment of pixels from each generated image and stack them horizontally (bottom). The desired result is hairs moving in animation, creating a time-varying field. With StyleGAN2 the hairs mostly stick to the same coordinates, creating horizontal streaks instead.

## 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 נייקויסט (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 פיצור 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.

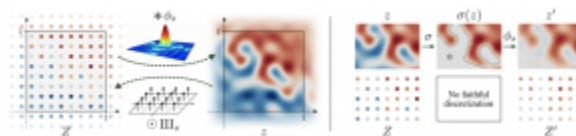


Figure 2: Left: Discrete representation  $Z$  and continuous representation  $z$  are related to each other via convolution with ideal interpolation filter  $\phi_s$  and pointwise multiplication with Dirac comb  $\text{III}_z$ . Right: Nonlinearity  $\sigma$ , ReLU in this example, may produce arbitrarily high frequencies in the continuous-domain  $\sigma(z)$ . Low-pass filtering via  $\phi_s$  is necessary to ensure that  $Z'$  captures the result.

## תקציר מאמר:

המאמר מציין שתי סיבות תופעת התקפלות תדרים המתוארת בפסקה הקודמת:

1. שימוש במסננים (פילטרים) כגון בילינארי או strided convolution במהלך יצירת תמונה.
2. 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  $3 \times 3$  ב-StyleGAN2 המקורי בקונבולוציות  $1 \times 1$  סימטריות שבאופן די ברור equivariant לסיבוב (נציין כי קונבולוציות  $3 \times 3$  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

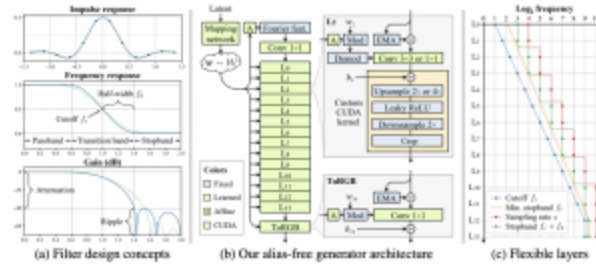


Figure 4: (a) 1D example of a  $2\times$  upsampling filter with  $n = 6$ ,  $s = 2$ ,  $f_c = 1$ , and  $f_h = 0.4$  (blue). Setting  $f_h = 0.6$  makes the transition band wider (green), which reduces the unwanted stopband ripple and thus leads to stronger attenuation. (b) Our alias-free generator, corresponding to configs  $\tau$  and  $\kappa$  in Figure 3. The main datapath consists of Fourier features and normalization (Section 3.1), modulated convolutions [33], and filtered nonlinearities (Section 3.2). (c) Flexible layer specifications (config  $\tau$ ) with  $N = 14$  and  $s_N = 1024$ . Cutoff  $f_c$  (blue) and minimum acceptable stopband frequency  $f_h$  (orange) obey geometric progression over the layers; sampling rate  $s$  (red) and actual stopband  $f_c + f_h$  (green) are computed according to our design constraints.

## איך מודדים equivariance להזהה ולסיבוב:

כדי לשערך את מידת ה-equivariance של התמונה  $\mathbf{X}$ , הנוצרת באמצעות מוקטור לטנטי  $\mathbf{w}$ , להזהה/סיבוב, יש להבין איזו טרנספורמציה צריך לעשות ל- $\mathbf{w}$ , 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  $\mathbf{w}$  שנבנה מוקטור לטנטי  $\mathbf{z}$  מהתפלגות גאוסית באמצעות העברתו של  $\mathbf{z}$  דרך רשת מיפוי (mapping network). תת-וקטורים של  $\mathbf{w}$  "מוזרקים" לשכבות שונות של רשת ה-synthesis ליצירת פיצ'רים בסקאלות שונות.
- וקטור דטרמיניסטי נלמד  $\mathbf{w}_0$ .

אז כדי להקל על חיפוש טרנספורמציה של  $\mathbf{w}_0$  שתגרום להזהה/סיבוב של תמונה בהיסט/זווית נתונה  $t$ , המחברים מציעים להחליף את  $\mathbf{w}_0$  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  $4\times 4$ ). החלפה זו מאפשרת למצוא את הטרנספורמציה  $\mathbf{T}$  ל- $\mathbf{w}_0$  כדי שהגנרטור ייצור מ- $\mathbf{T}(\mathbf{w}_0)$  את התמונה המוזזת/המסובבת, בצורה קלה.

לבסוף השערך של equivariance של התמונה  $\mathbf{X}$  להזהה/סיבוב נמדד באמצעות (peak signal-to-noise ratio (PCINR בין  $\mathbf{X}$  לתמונה המוזזת/המסובבת "המושלמת" לבין  $\mathbf{z}$ , הנוצרת מהווקטור הלטנטי המוזז/המסובב. לבסוף equivariance מחושב בתור ממוצע של PCINR-ים מעל סיבובים/הזזות של ווקטורי  $\mathbf{w}$  האפשריים.

בנוסף המאמר מציע כמה שינויים 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.

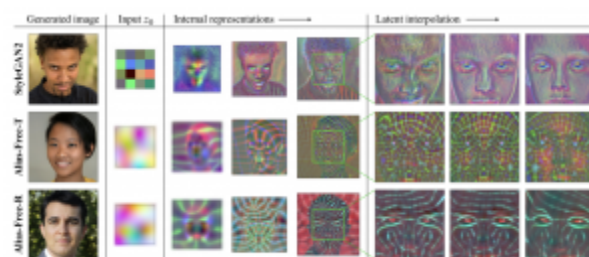


Figure 6: Example internal representations (3 feature maps as RGB) in StyleGAN2 and our generators.

## 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.

#deepnightlearners

הפוסט נכתב על ידי [מיכאל \(מייק\) ארליכסון](#), [PhD](#), Michael Erlihson.

מיכאל עובד בחברת הסייבר [Salt Security](#) בתור Principal Data Scientist. מיכאל חוקר ופועל בתחום הלמידה העמוקה, ולצד זאת מרצה ומנגיש את החומרים המדעיים לקהל הרחב.

ברצוני להודות עמוקות ל**ישי טובי**, מנכל טכנולוגי ב [EntityMed.com](#) על עזרתו בפיענוח עיבוד האותות במאמר. ותודה ענקית ל**לירון יצחקי**, ראש צוות מחקר [HourOne.ai](#) על סיעור המוחין בהבנת המאמר. ללא עזרתם של ישי ושל לירון סקירה זו לא היתה יוצאת לאור!!