קירה 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 מבית אמזון בתחום הלמידה העמוקה. המאמר הנסקר היום:

# GAN-Control: Explicitly Controllable GANs

## פינת הסוקר:

המלצת קריאה ממייק: חובה לאוהבי גאנים.

בהירות כתיבה: טובה מאוד.

רמת היכרות עם כלים מתמטיים וטכניקות של **ML/DL** הנדרשים להבנת מאמר: נדרשת הבנה טובה בארכיטקטורות עכשוויות של הגאנים (StyleGAN2) וידע בסיסי בנושא אימון של הגאנים. בנוסף נדרשת הבנה בסיסית של עקרונות הלמידה המנוגדת.

יישומים פרקטיים אפשריים: יצירה של תמונות פוטוריאליסטיות בעלות מכלול מוגדר של פיצ'רים ויזואליים כגון גיל, תנוחת ראש, צבע שיער וכדומה בכמה דומיינים כמו תמונות פנים מצוירות ותמונות פרצופים של חיות.

---

## פרטי מאמר:

לינק למאמר: זמין כאן

לינק לקוד: **לא שותף בארקיב**

פורסם בתאריך: 07.01.21, בארקיב

הוצג בכנס **:לא ידוע**

---

## תחומי מאמר:

- גאנים (GANs).

---

## כלים מתמטיים, טכניקות, מושגים וסימונים:

- למידה מנוגדת ((contrastive learning
- אימון של גאן עם פיצ'רים מופרדים בצורה מפורשת.

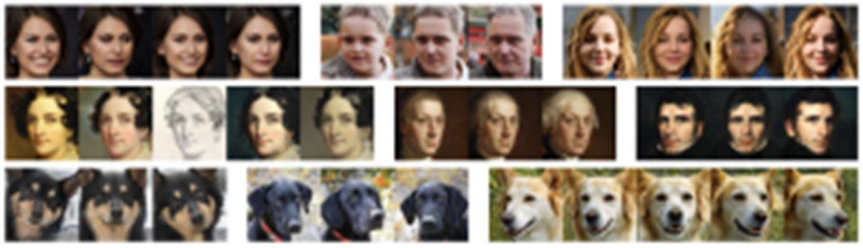

Figure 1: We propose a framework for training GANs in a disentangled manner which allows for explicit control over generation attributes. Our method is applicable to diverse controls in various domains. First row (left to right) demonstrates our control over facial expression, age and illumination of human portraits. Second row (left to right) demonstrates our control over artistic style, age and pose of paintings. Third row demonstrates our pose control over faces of dogs.

## מבוא והסבר כללי על תחום המאמר:

ליצירה של תמונות פוטוריאליסטיות בהינתן פרמטרים ויזואלים (כמו גיל, צבע שער, תאורה  וכדומה) באיכות גבוהה יש שימושים רבים במגוון תחומים כגון עיצוב גרפי, משחקי וידאו, קולנוע, תחום הצילומים הרפואיים ועוד. בשנים האחרונות נרשמו כמה בפריצות דרך בתחום הזה במיוחד בפיתוח מודלים יצירת צילומי פנים (face images) בעלות מכלול מוגדר של פיצ'רים ויזואליים. מודלים אלו בדרך כלל משתמשים בשיטות מידול D3 וסובלים לרוב מעלויות יצירה גבוהות ובעלות שונות נמוכה (יוצרות תמונות דומות אחת לשנייה). מצד שני מודלי גאנים עכשווים כמו StyleGAN2 מפגינים יכולת מרשימה ביצירת תמונות פוטוריאליסטיות באיכות מאוד גבוהה ובעלות יצירה סבירה אך מתקשות ליצור תמונות בעלות פיצ'רים ויזואליים נתונים. יש עבודות המשלבות את שתי הגישות האלו ומצליחות ליצור תמונות פנים פוטוריאליסטיות באיכות מרשימה בעלות תכונות נשלטות כמו תנוחה, תאורה וסוג הבעת פנים. אך מכיוון שמודלים אלו מתבססים על מידול D3 הם לא מאפשרים ליצור, למשל, תמונה של אדם בגיל מסוים כי מודלי D3 אינם מאפשרים זאת. בנוסף קשה להרחיב שיטות אלו לדומיינים קרובים כמו תמונות מצוירות או צילומים של פרצופי חיות אם אין ברשותנו מודלים של D3 המתאימים לתחומים אלו.

[scroll_highlight]המאמר הנסקר מציע גישה, הנקראת GAN-Control, הנותנת מענה לחולשה זה ומציע שיטה המאפשרת ליצור תמונות בעלות תכונות ויזואליות מוגדרות בצורה מפורשת.[/scroll_highlight] השיטה מצליחה ליצור תמונות פוטוריאליסטיות ב 3 דומיינים שונים: צילומי פנים, תמונות מצוירות וצילומי פרצופים של חיות. הגישה שלהם מאפשרת ליצור תמונות מגוונות לגיל, תאורה, תנוחה וצבע שער נתונים בצורה מפורשת לתחומים הנ"ל.

השיטה המוצעת מורכבת משני שלבים עיקריים:

- אימון של גאן עם פיצ'רים מופרדים בצורה מפורשת(explicitly disentangled features): בגדול (ההסבר המפורט יינתן בפרק הבא) מחלקים את המרחב הלטנטי הראשון ( **Z**ראה הערה על המרחבים הלטנטיים למטה) לתת-מרחבים **Z_i**כאשר כל תת מרחב אחראי על פיצ'ר ויזואלי ספציפי של התמונה המגונרטת (תת-המרחב האחרון אינו אחראי על פיצ'ר ספציפי ואחראי על שאר הפיצ'רים הלא נשלטים של תמונה)

- אימון של רשת הממפה פיצ'רים ויזואליים נתונים (בצורה מפורשת כמו למשל גיל או זווית צילום של תמונת פנים) לתת-מרחב הלטנטי "שלה". כתוצאה מכך תת מרחב **Z_age** , למשל, האחראי על הגילאים יחולק ל"איזורים" כך שכל "איזור" אחראי על גיל מסוים. שלב זה יאפשר לגנרט תמונה בעלת פיצ'רים ויזואליים נתונים:

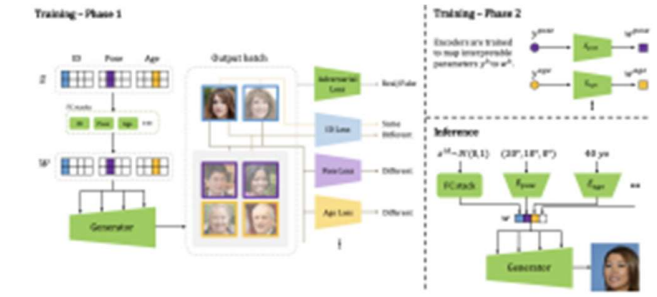


Figure 2: **Explicitly controllable GAN**: In Phase 1, we construct every batch so that for each attribute, there is a pair of latent vectors sharing a corresponding sub-vector, $z^k$. In addition to the adversarial loss, each image in the batch is compared in a contrastive manner, attribute-by-attribute, to all others, taking into account if it has the same or a different sub-vector. In Phase 2, encoders are trained to map interpretable parameters to suitable latent vectors. Inference: An explicit control over the attribute $k$ is achieved by setting the $E_k$'s input to a required value.

הערה לגבי המרחבים הלטנטיים של **GAN-Control**: המאמר משתמש בארכיטקטורה של 2StyleGAN שיש לו שני מרחבים לטנטיים **Z** ו- **W** כאשר הראשון הינו מרחב הלטנטי "הגולמי" הסטנדרטי של הגאנים והשני **W** הינו מרחב הסגנון הלטנטי עם פיצ'רים מופרדים האחראים על אספקטים ויזואליים שונים של התמונה המגונרטת). המיפוי מ- **Z** ל- **W** ממודל עי" רשת MLP בעלת 8 שכבות(עם משקלים נלמדים כמובן). המאמר מציע למדל כל מיפוי מתת-מרחבים **Z_i** ל **W_i** עי" רשת MLP שלה.

## הסבר של רעיונות בסיסיים:

נדון כעת איך מאמנים את GAN-Control לגנרט תמונות מפיצ'רים מופרדים.

שלב אימון 1: המאמר מציע להשתמש בלוס המנוגד לכל פיצ'ר. בשביל כך כל באטץ' נבנה מזוגות של וקטורים לטנטיים **z_b** המכילים תת-וקטור משותף ( **z_bk** המתאים לתת-מרחב הלטנטי **Z_k**) כאשר שאר התת-וקטורים הינם שונים. נציין שעבור זוג דוגמאות כאלו, הגנרטור של GAN-Control אמור ליצור שתי תמונות דומות בפיצ'ר ה-k בלבד ושונות בשאר הפיצ'רים. בשביל לבנות הלוס המנוגד המאמר מגדיר מרחק בין זוג תמונות I_1 ו- I_2 מבחינת הפיצ'ר k, המסומן 2_D_k(I_1, I) המודד דמיון בין התמונות בפיצ'ר k. למשל עבור הפיצ'ר שהוא זהות של אדם בתמונה D_k מודד "עד כמה I_1 ו- I_2 מתארות אותו אדם).

אז איך בעצם מחשבים את הלוס המנוגד כאן?

- עבור זוג תמונות עם פיצ'ר k זהה, אנו מנסים להגדיל את הדמיון בין התמונות הנוצרות מבחינת פיצ'ר k. כלומר המטרה הינה להקטין את המרחק ביניהן, הניתן עי" D_k מוגדר כמקסימום בין הפרש של 2_D_k(I_1, I) וקבוע c_k , לבין אפס. כלומר אנו "שואפים" שהמרחק המקסימלי בין I_1 ו- I_2 מבחינת פיצ'ר k יהיה c_k לכל היותר.
- עבור זוג תמונות עם פיצ'ר שונה j, המטרה למקסם את המרחק בין התמונות מבחינת פיצ'ר זה. הלוס במקרה הזה מוגדר עי" המקסימום בין אפס לבין ההפרש בין קבוע c_j* ו- 2_D_j(I_1, I). בדומה לסעיף הקודם המטרה כאן לגרום למרחק D_j להיות לכל הפחות c_j*.
- פונקציית לוס מוגדרת כסכום של הלוסים על כל הפיצ'רים

בסוף, כמקובל בגאן, מוסיפים ללוס המנוגד את הלוס האדברסריאלי של 2StyleGAN.

חישוב מרחקי D_k: בשביל למדוד מרחק בין תמונות מבחינת פיצ'ר k המאמר משתמש בייצוג במימד נמוך M_k של תמונה הנבנה עי״ רשת המאומנת למטרת זיהוי פיצ'ר k. למשל עבור דמיון בין תמונות מבחינת גיל, המאמר משתמש בייצוג מימד נמוך שנבנה עי״ הרשת לזיהוי גיל, למדידת דמיון בין תמונות מבחינת זהות של אדם המצולם, לוקחים את הייצוג מהרשת לזיהוי פנים (face detection). המאמר משתמש במרחקי L, 2L1 ומרחק cosine למדידת מרחק בין וקטורי ייצוג של תמונות מבחינת פיצ'רים ויזואליים שונים.

שלב אימון 2: המאמר מציע לבנות מיפוי(לאמן רשת) נפרד לכל פיצ'ר k ויזואלי כמו גיל (20, 30, 40 ,...) או זווית צילום (0, 5, 15,...) למרחב ״סגנון״ לטנטי שלו **W_k**. אבל איך מאמנים את הרשתות אלו? המאמר מציע את דרך אלגנטית לעשות זאת:

- מרגילים מספר וקטורי **z.**
- ממפים אותם למרחב ״סגנון״ **W**(זוכרים שכבר אימנו את המיפוי הזה בשלב הראשון).
- מגנרטים תמונות מוקטורי הסגנון **w**ומעבירים את התמונות דרך הרשתות לזיהוי כל פיצ'ר y_k.
- בונים דאטהסט מזוגות (**w_k**, y_k) לכל פיצ'ר k.
- מאמנים רשתות מקודדות E_k הממפות y_k ל- **w_k** (גם לכל פיצ'ר בנפרד).

אז מה קורה בזמן האינפרנס? זה מאוד פשוט - מזינים פיצ'רים ויזואליים y_k לרשתות מקודדות E_k ובונים וקטורי הסגנון **w_k.**בסוף משתמשים בגנרטור המאומן בשביל ליצור תמונה.

## הישגי מאמר:

המחברים השוו את התמונות הנבנות עם GAN-Control עם אלו שנוצרו עם השיטותDFG ו-CONFIG (שיוצרות תמונות פנים עם פרמטרים ויזואליים נשלטים ומבוססות על גישות מידול D3). המאמר הצליח להוכיח את עליונותה של GAN-Control על שיטות אלו ב- 3 דומיינים שונים בשני היבטים הבאים:

- מרחק inception של פרשה (FID) משופר(נמוך יותר) המצביע בדרך כלל על תמונות יותר פוטוריאליסטיות.
- דיוק משופר מבחינת התאמה של תמונה לפרמטרים הוויזואליים שאיתם היא נבנתה. למשל עבור תמונה הנוצרת עם זווית צילום של 30 מעלות, GAN-Control הצליח ליצור תמונות עם זוויות צילום קרובה יותר ל 30 מעלות בממוצע ועם שונות נמוכה יותר (זווית צילום של תמונה נמדדת עי״ רשת ייעודית מאומנת למשימה זו). דיוק מבחינת פרמטרים אחרים נמדד בצורה דומה.

[gallery size="medium" columns="2" link="file" ids="6769,6768"]

## דאטהסטים:AFHQ, MetFaces, FFHQ,

**נ.ב.** המאמר מציע גישה מאוד יעילה ואינטואיטיבית ליצירת תמונות בעלות פיצ'רים ויזואליים נשלטים באיכות גבוהה יותר מהשיטות המתחרות. הגישה מסוגלת ליצור תמונות פוטוראליסטית בעלות תכונות ויזואליות נתונות ב 3 דומיינים שונים: צילומי פנים, צילומי פרצופי חיות ותמונות מצוירות. הקוד לא שותף כרגע (אני מניח שהמחברים טרם הספיקו למלא בקשת פטנט - זה חשוב לחברה כמו אמזון). אני די בטוח הקוד יפורסם ממש קרוב. אני גם מחכה לראות את השימושים של גישה זו בדומיינים נוספים.

#deepnightlearners