# Review 40: Unsupervised Learning of Visual Features by Contrasting Cluster Assignments

**Paper: https://arxiv.org/abs/2006.09882v5**

פינת הסוקר:

המלצת קריאה ממייק: חובה בטח לאוהבי למידת הייצוג.

בהירות כתיבה: בינונית פלוס.

רמת היכרות עם כלים מתמטיים וטכניקות של ML/DL הנדרשים להבנת מאמר: הבנה טובה בעקרונות הלוס המנוגד וידע טוב באופטימיזציה.

יישומים פרקטיים אפשריים: למידה ייצוגים חזקים על דאטהסטים לא מתויגים עם תקציב חישוב מצומצם.

פרטי מאמר:

פורסם בתאריך: 08.01.21, בארקיב.

הוצג בכנס: NeurIPS 2020.

תחומי מאמר:

למידת ייצוג ללא דאטהסט מתויג (SSRL - self-supervised representation learning).

SSRL מבוססת על טכניקות קליסטור (Clustering for deep representation learning).

כלים מתמטיים, מושגים וסימונים:

מולטי-קרופ - טכניקת אוגמנטציה המבוססת על לקיחת פאטצ'ים קטנים של תמונה ברזולוציות נמוכות שונות.

האלגוריתם של סינקהורן קנופ (Sinkhorn-Knopp) לפתרון בעיית הטרנספורט האופטימלי למידות הסתברות דיסקרטיות.

תמצית מאמר:

המאמר מציע שיטת למידת ייצוג על דאטהסט לא מתויג. רוב גישות המודרניות בתחום הזה (SSRL) מורכבות משני מרכיבים עיקריים:

הלוס המנוגד (contrastive loss - CL): מסתמך על ההנחה שייצוגים של דוגמאות קרובות צריכים להיות קרובים, בזמן שייצוגים של דוגמאות לא קשורות (נבחרות רנדומלית בד"כ) צריכים להיות רחוקים.

שיטה ליצירה של דוגמאות "דומות", קרי אוגמנטציה: בדרך כלל זוג דוגמאות קרובות (אקרא לזוגות האלו בהמשך זוגות חיוביים או זוגות קרובים) נוצר עי" ההפעלה של שתי אוגמנטציות שונות על אותה דוגמא.

נציין כי גישות SSRL המודרניות מסתמכות של השוואה של מספר גבוה מאוד של זוגות ייצוגים של דוגמאות שמצריך כמות גדולה של זכרון ומשאבי עיבוד משמעותיים. דרישות אלו מקשות על יישום של שיטות אלו בצורת אונליין (לטענת המאמר רוב שיטות SSRL היום מיושמות בצורת אונליין שדי הפתיע אותי). אז בואו נדבר על החידושים שהמאמר הזה מציע:

שיטת אימון SwaV: המאמר הנסקר מציע שיטה חדשה SSLR (הנקראת SwaV) העשויה להוריד גם את כמות החישובים וגם לצמצם את כמות הזכרון הנדרשות. הרעיון העיקרי של המאמר הינו שינוי ״ההגדרה של מושג הדמיון בין ייצוגי דוגמאות״. למעשה המאמר ״מאלץ״ זוגות של הדוגמאות הקרובים ״להשתייך״ לאותם הקלאסטרים במרחב הייצוג במקום להשוות את הייצוגים בצורה מפורשת (שיוך זה המיוצג עי״ הקוד של דוגמא המחושב על סמך הבאטץ' שלו - אופן בנייתו יפורט בהמשך). נציין ש- SwaV אינו דורש לשמור בנק של דוגמאות שליליות שהופך אותו למועמד טוב למימוש בצורת אונליין.

שיטת אוגמנטציה מולטי-קרופ: המאמר מציע שיטת אוגמנטציה הנקראת מולטי-קרופ שמתחילה מהחישוב של שני ״קרופים סטנדרטיים״ x\_cr1 ו- x\_cr2 של תמונה x. לאחר מכן לוקחים ״קרופים קטנים יותר״ של x\_cr1 ו- x\_cr2 במגוון רזולוציות נמוכות ובונים מהם סט דוגמאות חיוביות עבור תמונה x. לטענת המאמר שיטה זו מקטינה את כמות החישובים הנדרשת תוך שמירה על הביצועים.

הסבר של רעיונות בסיסיים:

עכשיו ננסה להבין מה פונקצית המטרה L שבליבה של שיטת SwaV. פונקציית L מוגדרת באופן הבא (לכל דוגמא בבאטץ'):

בונים מספר אוגמנטציות לדוגמא x עם מולטי קרופ או כל גישה אחרת.

מרכיבים מאוגמנטציות אלו זוגות של דוגמאות.

בונים וקטורי ייצוג z\_i לכל הדוגמאות שבנינו.

לכל זוג וקטורי ייצוג (z\_1, z\_2) מחשבים את הקודים שלהם q\_1 ו- q\_2.

מחשבים את סכום הדמיונות l\_s בין z\_1 ו- q\_2 ובין z\_2 ל q\_1.

מחשבים את הסכום L\_x של l\_s של כל הזוגות של הדוגמאות החיוביות של דוגמא x.

פינת האינטואיציה:

למעשה תהליך אימון זה ״מאלץ״ וקטורי ייצוגי של דוגמא להכיל מידע על הקוד של הדוגמאות הקרובות. בצורה לא פורמלית ניתן לומר שאנו מנסים למקסם את ״המידע הדדי״ בין הייצוגים של הדוגמאות שזה המטרה העיקרית של האימון עם הלוס המונגד CL. דרך אגב השם של השיטה נובע מהפעולה שחלוף (swap) שמבצעים בין הייצוגים ובין הקודים של דוגמאות קרובות באימון.

השאלה האחרונה שטרם התייחסנו אליה הינה מבנה של פונקציית לוס בין ייצוג z לקוד q?

מבנה של פונקציה לוס בין ייצוג z לקוד q (של דוגמאות קרובות): אם אתם זוכרים הקוד q ניתן לפרש כווקטור הסתברויות שיוך לקלסטרים. למעשה אנו רוצים שהקוד q ישקף בצורה כמה שיותר טובה את המרחקים של z מהפרוטוטייפים c\_i שניתן לראות אותם בתור מרכזים (סנטרואידים) של קלסטרים של ייצוגים. אז קודם כל אנו בונים את וקטור המרחקים המנורמלים מ- z לכל c\_i. מרחק זה מחושב כאקספוננט של המכפלה הפנימית בין z ל c\_i. בסוף לוקחים את וקטור המרחקים ומנרמלים אותו. לאחר מכן מחשבים את קרוס אנטרופי בין q לוקטור מרחקים מנורמל שחישבנו. את הפונקציה זו אנו ממקסמים ביחס ל ייצוגים z וביחס לפרוטוטייפים c.

פינת האינטואיציה:

שימו לב על הדמיון של המרחק בין וקטור הייצוג z ל- c\_i לביטוי של החוב המנוגד CL. וזה לא מקרי - אתם זוכרים שלהבדיל משיטות מבוססות CL קלאסי, אין לנו כאן דוגמאות שליליות בצורה מפורשת. אז מה שמשחק כאן את תפקיד "הדוגמאות השליליות" זה מרכזי הקלסטרים שרחוקים מ z. כלומר הם מאלצים ייצוגים של דוגמאות חיוביות להיות רחוקים בצורה כמה שיותר דומה מכל הקלסטרים השליליים וקרובים באותה מידה מהקלסטרים החיוביים. לדעתי זה הנקודה הכי משמעותית במאמר (!!).

הסבר על בניית קוד q של ייצוג z: הקוד q של וקטור ייצוג z מתאר את ״רמת קרבתו״ של z ל K וקטורי פרוטוטייפ c\_i. וקטור c\_i "מייצג" את הקלסטר i. קוד של דוגמא (וגם של כל האוגמנטציות שלה) מחושב על סמך באטץ' בודד בלבד(!!). אפשר להגיד שהקוד q מייצג את ההסתברויות שיוך של וקטור הייצוג z של הדוגמא נתונה לקלסטרים המיוצגים עי״ וקטורי c\_k.

מטריצה Q המכילה את הקודים של כל הדוגמאות מהבאטץי הינה פתרון של בעיית אופטימיזציה לינארית עם איבר רגולריזציה השווה לאנטרופיה הכוללת של Q (עם מקדם קטן). פונקציה מטרה זו מנסה למקסם את הדמיון הכולל בין וקטורי ייצוג של הדוגמאות בבאטץי לפרוטוטייפים c\_i (כלומר לפזר את הקודים בצורה המשקפת את את יחס המרחקים בין ייצוג הדוגמא למרכזי הקלסטרים השונים). שימו לב שבעיית אופטימיזציה זו מזכירה בצורתה את בעיית הטרנספורט האופטימלי בין מידות הסתברות דיסקרטיות (האחידות) המוגדרות על שני דאטהסטים. את התפקיד של דאטהסטים כאן משחקים הפרוטוטייפים c\_i וקטורי הייצוג z\_i של כל הדוגמאות בבאטץ'. המטרה כאן זה למצוא את האופן האופטימלי שבו ניתן "להעביר את המסה ההסתברותית מווקטורי z\_i לוקטורי c\_i (נציין שפונקצית המרחק שיש בהגדרה של הטרנספורט האופטימלי הינה פרופורציונלית במקרה שלנו למרחק בין z ל- c). למעשה אנו מנסים למצוא מטריצה Q האי שלילית, שאיבר (j,k) שלה מגדיר את המסה ההסתברותית המועברות מווקטור z\_k לוקטור c\_j, כלומר הסתברות השיוך של z\_k לקלסטר של c\_j. מכיוון שאנו רוצים שאותו מספר דוגמאות "ישויך" לכל קלסטר, מוסיפים אילוץ על סכום השורות וסכום העמודת של Q. בעיה זו פותרים בעזרת אלגוריתם איטרטיבי של סינקהורן-קנופ.

הסבר על מושגים חשובים במאמר:

שיטות אימון של גישות SSRL המודרניות: בדרך כלל בזמן האימון של SSLR לכל זוג של דוגמאות קרובות בונים מספר גדול של זוגות רנדומליים (אקרא לזוגות כאלו זוגות רחוקים או זוגות שליליים). כאן פונקציית המטרה F\_ob (שממקסמים אותה) הינה יחס בין אקספוננט של דמיון של ״הזוג הקרוב״ (בין הייצוגים שלהם) לסכום הדמיונות בינו לבין כל הזוגות שליליים. למשל בשיטת SimCLR כל באטצי מורכב מ-N זוגות של דוגמאות קרובות (אוגמנטציה של אותה הדוגמא) המהווים את הזוגות החייבים כאשר עבור דוגמא נתונה, כל הדוגמאות פרט ל״בת הזוג״ שלה נחשבת לדוגמא שלילית עבורה. פונקציה המטרה לכל באטץי הינה סכום של פונקציות המטרה של כל 2N דוגמאות של הבאטץ'.

בנק של ייצוגי דוגמאות שליליות: ידוע שהגדלת מספר הזוגות השליליים לכל זוג חיובי באימון תורמת לעוצמת הייצוג של הדאטה. כתוצאה מכך משתמשים בבאטצ'ם מאוד גדולים (עשרות אלפי דוגמאות) שדורש משאבי זכרון גדולים, כח עיבוד רב (צריך לחשב את הייצוג של עשרות אלפי דוגמאות מהבאטץי). כדי להקטין את כוח העיבוד הנדרש הוצע (MOCO) ״בנק הדוגמאות השליליות״ מהבאטצים הקודמים המכיל את הייצוגים של הדוגמאות מכמה הבאטצים הקודמים. כל פעם דוגמים משם ייצוגים של דוגמאות שליליות ומוסיפים את זה לייצוגים השליליים מהבאטץי הנוכחי. צריך לזכור שגישה זו כרוכה בהקצאת משאבי אחסון נוספים לשמירת בנק זה.

הישגי מאמר:

המאמר מראה ש-SwaV משולב עם מולטי-קרופ מצליח לייצר ייצוגים יותר חזקים משיטות בניית ייצוג רבות עבור מספר משימות. ההשוואה בוצעה בדרך הסטנדרטית: הוספה של שכבה לינארית לרשת הבונה ייצוג (עם משקלים מוקפאים) ובחינת ביצועיה על משימה מסוימת. קודם כל הם הראה שייצוג שנבנה באמצעות SwaV מציג ביצועים יותר טובים על דאטהסטים VOC07 ,iNaturalist2018 ו- Places205 מהייצוגים הנבנים על ImageNet מתויג (!!) גם על משימת סיווג ועל משימת זיהוי אובייקטים. בנוסף הם הראו שהייצוגים שלהם משיגים ביצועים יותר טובים מבחינת Top1/Top5 (לוקחים 1/5 דוגמאות הכי קרובות מבחינת הייצוג ומחשבים כמה מתוכם שייכים לאותה קטגוריה) מ- SimCLR ו- MoCov2. נזכיר שלהבדיל מ-MoCov2 , אין צורך בשמירה של בנק דוגמאות שליליות ב-SwaV. הם גם הראה את עליונותה של SwaV במשימות semi-supervised על שיטות כמו UDA ו- FixMatch. וזה רק חלק קטן מכל השוואות שהם עשו - הם באמת עשו עבודה מרשימה בהיבט הזה.

נ.ב.

מאמר ממש מגניב עם רעיון מסקרן המשלב תובנות רבות ממגוון שיטת SSRL. הם גם טרחו להשוות את הביצועים של השיטה שלהם מול מגוון רחב של אלגוריתמים, משימות, דאטה סטים וקונפיגורציות שזה בהחלט מרשים. בקיצור המלצת קריאה לוהטת ממני:)

שנקרא: