Medical Visual Question Answering

09/12/2018

Avi Turner - 039991708

[avi.turner111@gmail.com](mailto:avi.turner111@gmail.com)

055-6699069

תוכן

[תקציר 3](#_Toc532235140)

[רקע 4](#_Toc532235141)

[למידה עמוקה (Deep learning) 4](#_Toc532235142)

[מנוחים נוספים 4](#_Toc532235143)

[VQA - Visual Question Answering 5](#_Toc532235144)

[מהות ותיאור הבעיה 6](#_Toc532235145)

[VQA-Med 6](#_Toc532235146)

[מהות ותיאור הפתרון 7](#_Toc532235147)

[תוצאות ראשוניות 8](#_Toc532235148)

[מדדים 8](#_Toc532235149)

[WBSS 8](#_Toc532235150)

[BLEU 8](#_Toc532235151)

[תוצאות מ2018: 8](#_Toc532235152)

[השוואה לספרות 9](#_Toc532235153)

[רשימת ספרות 10](#_Toc532235154)

[מקורות נוספים 11](#_Toc532235155)

# תקציר

Visual Question Answering או בקיצור – VQA, הוא תחום מחקר על בניית מערכות תכנה המסוגלות לענות על שאלות, המנוסחות בשפה טבעית, לגבי תמונה נתונה. כלומר, בהינתן תמונה ושאלה על התמונה (קלט) תופק תשובה מתאימה (פלט)

מערכת תכנה המסוגלת להבין תוכן של תמונה ולתאר אותה בשפה טבעית הוא תחום חדש שמקבל תשומת לב רבה בקהילות הראיה הממוחשבת ועיבוד שפות טבעיות.

בתחום זה נעשה שימוש נרחב ביסודות שמקורן בDeep Learning , עיבוד שפות טבעיות וראייה ממוחשבת.

התחום הוא חדש יחסית (המאמר המוקדם היודע לי הוא מ2014), ועדיין יש בו מקום למחקר והתקדמות רבה.

בפרויקט זה, נתמקד ביישום של מערכת שכזו בהקשר של מערכות רפואיות, ובפרט אבחונים רפואיים המתוארים בשפה טבעית על סמך תמונות ממכשירי דימות רפואיים ושאלות לגביהם המנוסחות גם כן בשפה טבעית.

אנו מצפים להשיג התקדמות מבחינת איכות התוצאות לעומת עבודות דומות שנעשו במסגרת תחרות ImageCLEF 2018

# רקע

## למידה עמוקה (Deep learning)

למידה עמוקה, או Deep Learning, (Litjens et al., 2017) היא רשת נוירונים בעלת מספר שכבות נסתרות, כלומר שאינן שכבות קלט או פלט. בפועל, כאשר מתייחסים לרשת למידה עמוקה הכוונה בדרך כלל למספר רב של שכבות. ההגדרה של "מספר רב" גדל עם התפתחות הטכנולוגיה. החל מ-8 ב2012 ועד למעלה מ-150 ב2015 (איור 2)

יסודות רשתות הנוירונים הקלאסיות נטועות כבר בשנת **1943** אז פורסם מאמר (McCulloch and Pitts) שהציע מודל פשוט לפעולת הנוירון, שהיווה בסיס להתפתחותו של מדע הרשתות הנוירונים. וב**1957** היה כבר יישום של רשת נוירונים ויחידת הבסיס –נוירון מלאכותי - פרספטרון.

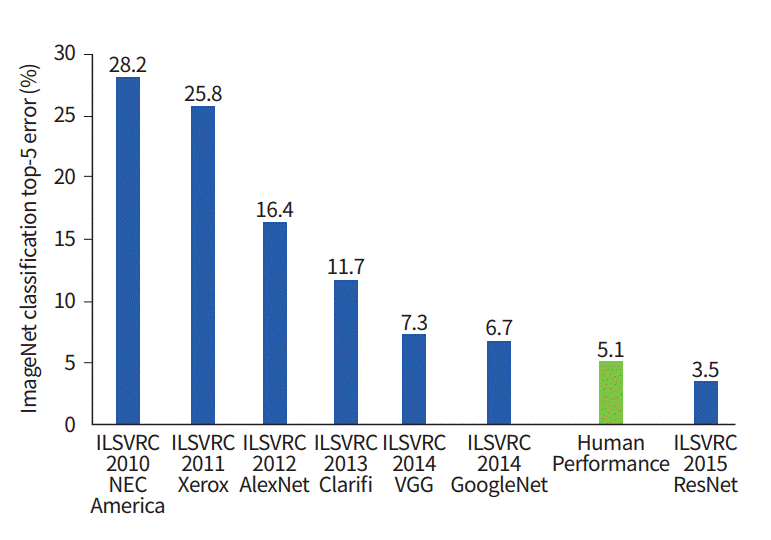
ב**1969** פורסם ספר שתקף את הרעיונות העומדים מאחורי הרשתות העצביות. הספר הציג פגם בנוירון הבסיסי. טיעונים שהוצגו בספר, ובעיקר הטיעון שלא משנה אילו משקולות נשים על הקלט, לא נוכל ליצור את השער הלוגי *xor*, עיכבו את המשך התפתחותן של הרשתות העצביות במשך שנים רבות.

ב**1980** "התגלה מחדש" אלגוריתם הלמידה שלמעשה פורסם כבר ב1974 לרשתות נוירונים הפך אותו לנפוץ ואפשר למידה של מודלים לא ליניאריים.

התקדמות נוספת חלה ב**1989** כאשר סביינקו פרסם הוכחה שרשת נוירונים יכולה לקרב כל פונקציה, וב**1990** כבר היו יישומים מוצלחים במגוון תחומים כמו רפואה, שיווק, ניהול סיכונים ועוד.

בתחילת שנות ה**2000**, במקביל לשיפור ניכר בביצועי החומרה, עידן האינטרנט ו"התפוצצות המידע" רשתות נוירונים החלו לתת תוצאות טובות בתחום הלמידה החישובית. עם המשך התקדמות החומרה ובפרט השימוש במעבדים גרפיים, המודלים של רשתות הנוירונים הפכו להיות בעלי יותר שכבות, ויכלו לפתור בעיות מורכבות יותר רשתות אלו נודעו בשם רשתות **למידה עמוקה**.

מאז **2012**, כל המנצחות בתחרויות גדולות בתחום זיהוי האובייקטים באמצעות ראייה ממוחשבת מבוססות על רשתות למידה עמוקה. ומאז 2015 – עם ביצועים יותר טובים משל בן אדם (איור 2)

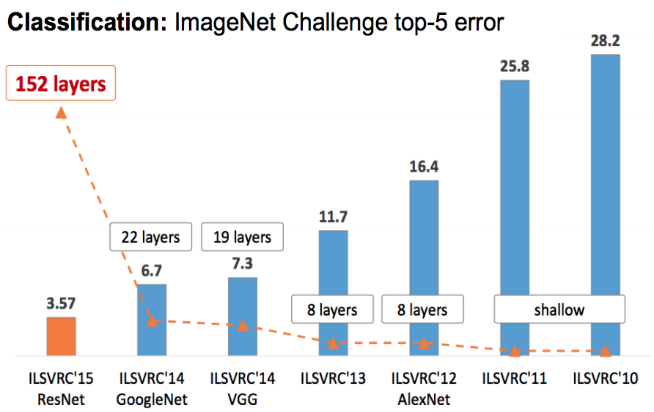


איור2

אחוזי הטעות של רשתות למידה עמוקות ביחס לביצועי בן אדם

איור 1

אחוזי הטעות של רשתות למידה עמוקות ומספר השכבות ברשת



## מנוחים נוספים

1. **NLP** – Natural language processing – עיבוד שפות טבעיות
2. **RNN** - Recurrent neural network - רשתות נוירונים בהן הפלט של כל שכבה משמש גם כקלט של אותה השכבה לצעד הבא. ארכיטקטורת היזון חוזר זאת מאפשרת לרשת 'לזכור' את המידע מהצעד הקודם (ושלמעשה הצטבר עד כה) ובכך לאפשר תיאור של רצפים. בשימוש נרחב בעיבוד שפות טבעיות.
3. **LSTM** – Long Short-Term Memory - גוון של RNN, נוצר כמענה לבעיה ברשתות RNN ֵשל גרדיאנט נעלם ששואף אסימפטוטית ל0/1 בזמן הלמידה (גרדיאנט נעלם / מתפוצץ)
4. **VGG**, **ResNet** – רשתות עמוקות (19 ו152 שכבות בהתאמה) (איור 1) מהרשתות שניצחו בתחרויות זיהוי אובייקטים ב2014 / 2015 בהתאמה (איור 2). שתיהן נמצאות גם כיום בשימוש נרחב בתחום הראיה החישובית.
5. **FC** – Fully Connected – שכבה ברשת נוירונים בה כל האיבר בשכבה מחוברים לכל האיברים בשכבה הבאה

## VQA - Visual Question Answering

עם התפתחות רשתות הלמידה העמוקה, הטכנולוגיה, ובפרט שימוש במעבדים גרפים, וזמינות המידע, התאפשרה למידה של מידע גולמי ומורכב יותר כמו שפה טבעית (NLP) ותמונות גולמיות. כלומר – כאשר הקלט הוא התמונה עצמה ולא מידע שהופק מהתמונה.

VQA הוא בעצם התפתחות טבעית של שני הענפים הנ"ל תוך שילובם למערכת אחת.

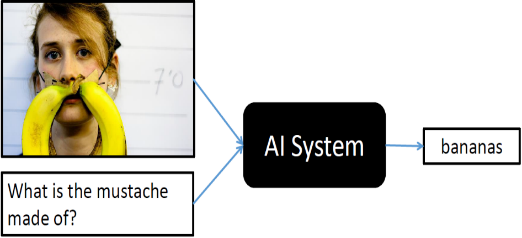
תחום מחקר זה עוסק בבניית מערכות תכנה המסוגלות לענות על שאלות, המנוסחות בשפה טבעית, לגבי תמונה נתונה.

התחום הוא חדש יחסית (ב2014 פורסם מאגר המידע והמאמר הראשון בנושא[[1]](#footnote-1)), ועדיין יש בו מקום למחקר והתקדמות רבה.

כיום, האלגוריתמים המובילים מגיעים לדיוק של [[2]](#footnote-2)70% לעומת דיוק של 83% אצל בני אדם – יש לציין שנתונים אלו הושגו על מאגר מידע בסדר גודל של מאות אלפי תמונות.

להמחשה:

בהינתן תמונה ושאלה, על המערכת לתת תשובה לשאלה:



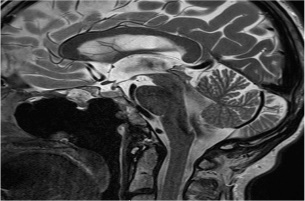
VQA

עבור תמונה מסוימת יכולות להיות מספר שאלות/תשובות. השאלות והתשובות מנוסחות בשפה טבעית.



# מהות ותיאור הבעיה

## VQA-Med



what does t2 weighted sagittal magnetic resonance scan show?

VQA

total excision tumor with packg material seen posterior part clivus extradural plane

מערכת תכנה המסוגלת להבין תוכן של תמונה ולתאר אותה בשפה טבעית הוא תחום חדש שמקבל תשומת לב רבה בקהילות הראיה הממוחשבת ועיבוד שפות טבעיות.

מערכת שכזו נותנת פתרון לכמה בעיות:

1. במסגרת עבודתם, מומחים וחוקרים רבים כותבים דו"חות על ממצאים מיקרוסקופים שהם מוצאים בתמונות רפואיות. מערכת אוטומטית שתתאר מה יש בתמונה תוכל להפחית משמעותית מהעומס המוטל עליהם ולשחרר צווארי בקבוק בכל תהליך השירות הרפואי.
2. כיום, במרכזים רפואיים מוטל עומס רב על רופאים, מצב שעלול, וגורם, לחוסר תשומת לב מספקת לכל אבחון – ועקב כך לסיכון חיי אדם. מערכת שכזו תוכל להוות מערך בקרה ו"להרים דגל" על כל אבחון ששונה מאבחון פלט של המערכת על מנת שבעל מקצוע יוכל לעבור רק על ההבדלים שהרופא והמערכת חלוקים לגביהם עם תשומת לב יתרה.
3. במקומות רבים, גם בעולם המערבי, רפואה מתקדמת אינה נחלתם של מעוטי יכולת. מערכת VQA-Med, בהנחה שתבשיל לכדי מערכת בשלה, תוכל לתת אבחון מהיר וזול למי שאין להם רפואה זמינה. גם בהיעדר בשלות כזו – היא תוכל לתת מענה לחוות דעת שניה זולה וכמעט אוטומטית למי שידם אינה משגת.

כאמור, התחום עדיין בחיתוליו וסביר להניח שבמסגרת פרויקט זה לא נייצר מערכת בשלה מספיק כדי לתת מענה מלא לבעיות הנ"ל, אבל כן יש ציפייה להניח אבן דרך נוספת בדרך ליעדים אלו.

ImageCleff היא מסגרת שמנוהלת בהתנדבות בעיקר ושמה לעצמה מטרה לקדם ולהעריך התקדמות בתחום העיבוד החזותי והרפואי בפרט.

החל מ2018, החלו תחרויות VQA שמתמקדות בתחום הרפואי – VQA Med

במסגרת פרויקט זה, אנו מצפים להשיג התקדמות מבחינת איכות התוצאות לעומת עבודות דומות שנעשו במסגרת תחרות ImageCLEF 2018

# מהות ותיאור הפתרון

בניית מערכת VQA המקבלת כקלט:

* תמונות
* צמדים של שאלות ותשובות על התמונות הנתונות

בדומה למודלים שפותחו בשנים האחרונות נשתמש במבנה הכללי הבא:

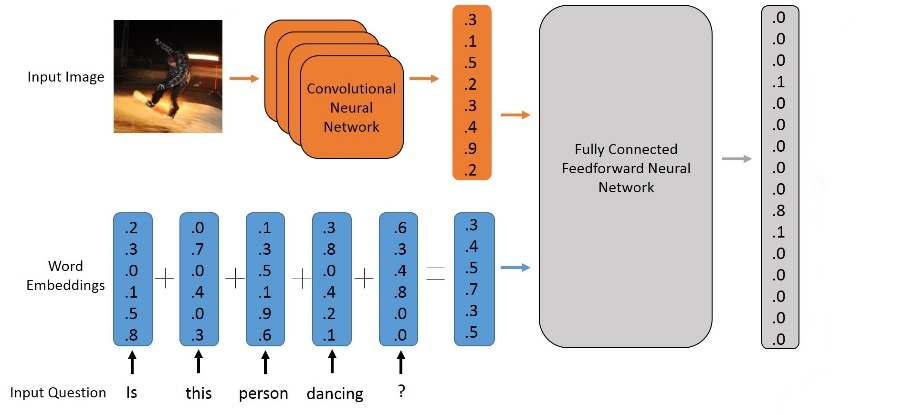
1. רשת המקבלת קלט דרך שני ענפים (איור 3) (Atg, 2018):

* ענף למידע החזותי. קלט: התמונה, פלט: וקטור המיצג את תכונות התמונה (a).
* ענף לטקסט של השאלה, פלט: וקטור המיצג את תכונות הטקסט (b).

ענף התמונה משתמש ברשת Deep Learning מאומנת (לדוגמא VGG) כדי לחלץ תכונות מהתמונה שמקודדות בצורה של וקטור.

ענף הטקסט משתמש ברשתות מסוגLSTM לוקח קידוד NLP של השאלה ואת המצב פנימי של LSTM כתכונות ומחלצת גם כן וקטור שמייצג תכונות אלו.

לאחר מכן שני הענפים משולבים באמצעות שכבתFC ורשת נוירונים קלאסית שהפלט שלה הוא ההסתברות להימצאות כל אחת מהמילים בקורפוס בתשובה לשאלה והתמונה הנתונה.



(b)

(a)

איור 3

מבנה קלסי של רשת VQA

הקלט למערכת הוא התמונות ושאלות / תשובות שמתקבלות במסגרת התחרות. בינתיים – השתמשנו בתמונות שמקורן בתחרות של 2018 שכוללות 2280 תמונות, שאליהן משויכות 5300 שאלות ותשובות. בשביל הערכת המערכת קיבלנו סט דומה בגודל של 325 תמונות ו500 שאלות/תשובות

מקור הקורפוס הוא באתגר קודם של ImageCLEF (ImageCLEF 2017 caption prediction), בו המשימה הייתה לחזות מה תהיה הכתובית המוצמדת לתמונה רפואית בכתבות מ PubMed Central (ארכיון פתוח של כתבות ומאמרים רפואיים).

על תמונות + כתוביות אלו, הופעל אלגוריתם של מחולל שאלות / תשובות. בשלב הראשון האלגוריתם מפשט משפטים מורכבים לכמה משפטים פשוטים, ואז מחליף את נושא המשפט במילת שאלה, וקובע את הנושא כתשובה לשאלה. לסיום, על השאלות / תשובות עוברו גם צמד מומחים שווידאו שהתוצאה סבירה.

* המידע נראה קצת "מלוכלך" - רבות מהשאלות והתשובות מנוסחות בצורה עילגת, ויש תמונות שנראות כלא שייכות לסט.

what shows air space consolidation in left lower lobe?

CT



what does 3d CT view show?

rotational movements measured

# תוצאות ראשוניות

## מדדים

### WBSS

איור 4

ויזואליזציה של חישוב עומק (הפונקציה d)

מדד זה מעריך את הפלט שמתקבל על בסיס הדמיון (similarity) הסמנטי בתחום הביו-רפואי.

WBSS מחשב את ניקוד הדמיון בין התשובה (הפלט) ובין "האמת המוחלטת" (ground truth) המסופקת לו כפרמטר.

הניקוד מתבסס על הדמיון הסמנטי ברמת המילה.

המילים בקורפוס, במקרה שלנו – קורפוס בעל אופי רפואי מסודר בצורה היררכית כגרף. דמיון בין שתי מילים מיוצג על ידי אורך הנתיב הקצר ביותר בין שתי מילים אלו בגרף. בהנחה שהגרף נבנה בצורה נכונה, מילים בעלות משמעות דומה יהיו בעלות נתיב קצר יותר ביניהן ממלים בעלות משמעות אחרת. למשל, [*גידול*, *סרטן*] לעומת [*גידול, שבר*] נצפה שהמרחק בין *גידול* ו*סרטן* יהיה קצר יותר.

הציון מתקבל על ידי:

*כאשר c1 ו c2 הן המילים שרוצים לחשב את הדמיון ביניהן, וc0 הוא המילה הספציפית ביותר הנמצאת בנתיב לשני המילים. כלומר האב הקדמון המשותף הקרוב ביותר לשתי המילים* (Buscaldi, Tournier, Aussenac-gilles, & Mothe, 2012)*.*

### BLEU

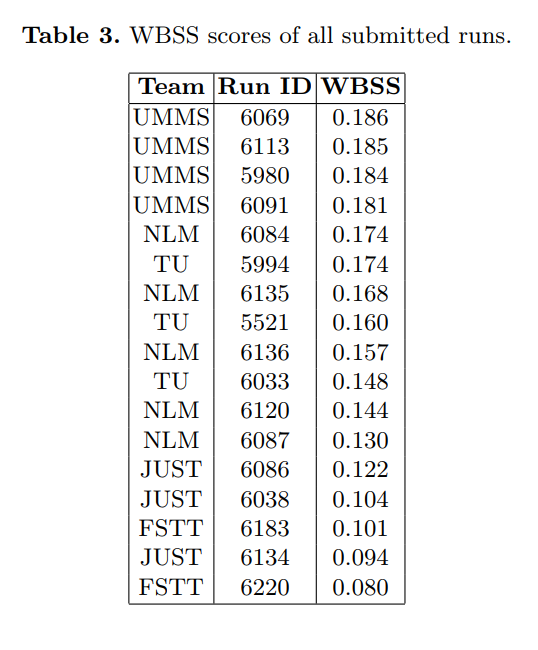
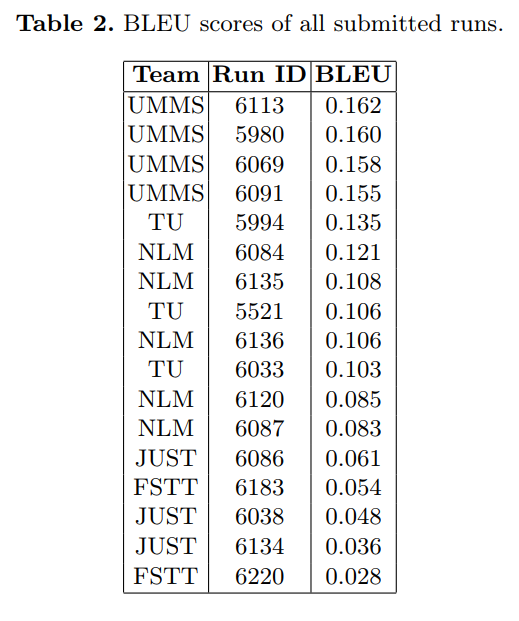
אלגוריתם להערכת איכות טקסט אשר תורגם ע"י מכונה תוך השוואה ל"אמת מוחלטת" (ground truth).

במהלך ההערכה:

* המילים עוברים tokenization - ממשפט לקורפוס של מילים:
  + כל האותיות מומרות לlower case
  + סימני הפיסוק מוסרים.
* מילות נפוצות (stop words) מוסרות.
* Stemming – ייצוג ההטיות של מילה ע"י ייצוג בודד (*fishing*, *fished*, *fisher*  *fish*)
* מופק ציון שהוא תוצר של מספר n-יות המילים המופיעות בתשובה ובאמת המוחלטת, אין שום חשיבות לנכונות או למבנה התחבירי של המשפט.

## תוצאות מ2018:

(Hasan et al., 2018)



תוצאות הגשה שביצענו לתחרות של 2018 (אחרי סיומה):

|  |  |  |
| --- | --- | --- |
|  | BLEU | WBSS |
| 19/10/2018 | 0.157 | 0.155 |
| 19/10/2018 | 0.146 | 0.143 |

# השוואה לספרות

מכיוון שהתחום עדיין בחיתוליו המקור היחיד שמצאתי שעוסק בVQA בכלל ובתחום הרפואי בפרט היה ImageCLEF, ארגון שמטרתו היא לקדם את תחום הראיה הממוחשבת ולמידה חישובית, ובמיוחד את השילוב עם טקסט בשפה טבעית.

מאגר התמונות-שאלות-תשובות שהשתמשנו בו הגיע התחרות של 2018, בו הזוכה הגיע לציון של:

|  |  |  |
| --- | --- | --- |
|  | BLEU | WBSS |
| בזמן תחרות | 0.158 | 0.186 |
| הגשה לאחר תחרות | 0.188 | 0.209 |

מסיכום התחרות עלה שבאופן כמעט גורף, רוב המתמודדים השתמשו במערכת כזו או אחרת של למידה עמוקה שכללה חלק או כל מהמאפיינים הבאים:

1. קידוד השאלה באמצעות ספריות / אלגוריתמי NLP
2. רשתות קונבולציוניות עמוקות לקידוד התמונה (עם או בלי רשתות מאומנות כמו VGG / ResNet)
3. Recurrent networks (RNN) לקידוד השאלה (עם או בלי מערכות מאומנות מקורפוס אחר)

חלק מהמתמודדים התמודדו עם הבעיה כבעיה של תיוג בה ניתן לבחור בתיוג אחד או יותר [[3]](#footnote-3) , בעוד אחרים ניסו לחולל תשובה[[4]](#footnote-4).

מאפיין נוסף שהיה אצל חלק מהמתמודדים היה מימוש Attention[[5]](#footnote-5) כדי לקשר בין המשפט לתמונה במבנה הרשת.

היו גם מספר ניסיונות לשימוש בכלים מתקדמים יותר כמו (Hasan et al., 2018):

1. Multi modal compact bilinear (MCB) pooling
2. multimodal factorized bilinear (MFB) pooling
3. [[6]](#footnote-6)Embedding based topic modeling (ETM)

בישום שלנו אנו גם כן משתמשים ב:

1. קידוד השאלה באמצעות רשת RNN, ובפרט – LSTM.
2. רשתות קונבולציוניות עמוקות שכבר אונו לקידוד התמונה.
3. קידוד מילות השאלה באמצעות ספריית NLP שאומנה על ויקפדיה
4. מתוכנן לבדיקה: שימוש בAttention

נציין כי בתחרויות של תרגום משפה לשפה, כאשר משתמשים במדד BLEU מתקבלים ציונים בסביבות 0.33, כאן הציונים היו נמוכים משמעותית. ואחד מהלקחים שעלו מהתחרות עולה כי קיים צורך להגדיל את גודל הקלט כדי לקבל תוצאות משמעותיות יותר.

# רשימת ספרות

Agrawal, A., Lu, J., Antol, S., Mitchell, M., Zitnick, C. L., Batra, D., & Parikh, D. (n.d.). VQA : Visual Question Answering, 1–23.

Atg, U. (2018). Read This Month, (January).

Buscaldi, D., Tournier, R., Aussenac-gilles, N., & Mothe, J. (2012). IRIT : Textual Similarity Combining Conceptual Similarity with an N-Gram Comparison Method. *First Joint Conference on Lexical and Computational Semantics (\*SEM 2012)*, 552–556.

Fukui, A., Park, D. H., Yang, D., Rohrbach, A., Darrell, T., & Rohrbach, M. (n.d.). for Visual Question Answering and Visual Grounding.

Hasan, S. A., Ling, Y., Farri, O., Liu, J., Henning, M., & Lungren, M. (2018). Overview of ImageCLEF 2018 Medical Domain Visual Question Answering Task.

Leonel, E. D. (2009). Phase transition in dynamical systems: Defining classes of universality for two-dimensional hamiltonian mappings via critical exponents. *Mathematical Problems in Engineering*, *2009*, 1–9. https://doi.org/10.1155/2009/367921

Lin, X., & Parikh, D. (n.d.). Active Learning for Visual Question Answering : An Empirical Study arXiv : 1711 . 01732v1 [ cs . CV ] 6 Nov 2017, 1–23.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., … Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*. https://doi.org/10.1016/j.media.2017.07.005

Zhang, P., Goyal, Y., Summers-stay, D., Batra, D., Tech, V., & Parikh, D. (n.d.). Yin and Yang: Balancing and Answering Binary Visual Questions.

Agrawal, A., Lu, J., Antol, S., Mitchell, M., Zitnick, C. L., Batra, D., & Parikh, D. (n.d.). VQA : Visual Question Answering, 1–23.

Atg, U. (2018). Read This Month, (January).

Buscaldi, D., Tournier, R., Aussenac-gilles, N., & Mothe, J. (2012). IRIT : Textual Similarity Combining Conceptual Similarity with an N-Gram Comparison Method. *First Joint Conference on Lexical and Computational Semantics (\*SEM 2012)*, 552–556.

Fukui, A., Park, D. H., Yang, D., Rohrbach, A., Darrell, T., & Rohrbach, M. (n.d.). for Visual Question Answering and Visual Grounding.

Hasan, S. A., Ling, Y., Farri, O., Liu, J., Henning, M., & Lungren, M. (2018). Overview of ImageCLEF 2018 Medical Domain Visual Question Answering Task.

Leonel, E. D. (2009). Phase transition in dynamical systems: Defining classes of universality for two-dimensional hamiltonian mappings via critical exponents. *Mathematical Problems in Engineering*, *2009*, 1–9. https://doi.org/10.1155/2009/367921

Lin, X., & Parikh, D. (n.d.). Active Learning for Visual Question Answering : An Empirical Study arXiv : 1711 . 01732v1 [ cs . CV ] 6 Nov 2017, 1–23.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., … Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*. https://doi.org/10.1016/j.media.2017.07.005

Teney, D., & Centre, A. (2017). Tips and Tricks for Visual Question Answering : Learnings from the 2017 Challenge.

Zhang, P., Goyal, Y., Summers-stay, D., Batra, D., Tech, V., & Parikh, D. (n.d.). Yin and Yang: Balancing and Answering Binary Visual Questions.

Agrawal, A., Lu, J., Antol, S., Mitchell, M., Zitnick, C. L., Batra, D., & Parikh, D. (n.d.). VQA : Visual Question Answering, 1–23.

Atg, U. (2018). Read This Month, (January).

Buscaldi, D., Tournier, R., Aussenac-gilles, N., & Mothe, J. (2012). IRIT : Textual Similarity Combining Conceptual Similarity with an N-Gram Comparison Method. *First Joint Conference on Lexical and Computational Semantics (\*SEM 2012)*, 552–556.

Fukui, A., Park, D. H., Yang, D., Rohrbach, A., Darrell, T., & Rohrbach, M. (n.d.). for Visual Question Answering and Visual Grounding.

Goyal, Y., Khot, T., Dhruv, D. S., & Devi, B. (n.d.). Making the V in VQA Matter : Elevating the Role of Image Understanding in Visual Question Answering.

Hasan, S. A., Ling, Y., Farri, O., Liu, J., Henning, M., & Lungren, M. (2018). Overview of ImageCLEF 2018 Medical Domain Visual Question Answering Task.

Leonel, E. D. (2009). Phase transition in dynamical systems: Defining classes of universality for two-dimensional hamiltonian mappings via critical exponents. *Mathematical Problems in Engineering*, *2009*, 1–9. https://doi.org/10.1155/2009/367921

Lin, X., & Parikh, D. (n.d.). Active Learning for Visual Question Answering : An Empirical Study arXiv : 1711 . 01732v1 [ cs . CV ] 6 Nov 2017, 1–23.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., … Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*. https://doi.org/10.1016/j.media.2017.07.005

Teney, D., & Centre, A. (2017). Tips and Tricks for Visual Question Answering : Learnings from the 2017 Challenge.

Zhang, P., Goyal, Y., Summers-stay, D., Batra, D., Tech, V., & Parikh, D. (n.d.). Yin and Yang: Balancing and Answering Binary Visual Questions.

## מקורות נוספים

(Agrawal et al., n.d.)

Agrawal, A., Lu, J., Antol, S., Mitchell, M., Zitnick, C. L., Batra, D., & Parikh, D. (n.d.). VQA : Visual Question Answering, 1–23.

(Lin & Parikh, n.d.)

Lin, X., & Parikh, D. (n.d.). Active Learning for Visual Question Answering : An Empirical Study arXiv : 1711 . 01732v1 [ cs . CV ] 6 Nov 2017, 1–23.

(Zhang et al., n.d.)

(Teney & Centre, 2017)

(Goyal, Khot, Dhruv, & Devi, n.d.)

Goyal, Y., Khot, T., Dhruv, D. S., & Devi, B. (n.d.). Making the V in VQA Matter : Elevating the Role of Image Understanding in Visual Question Answering.

(Gupta, n.d.)

Gupta, A. K. (n.d.). Survey of Visual Question Answering: Datasets and Techniques.

Deep Learning and Visual Question Answering

<https://towardsdatascience.com/deep-learning-and-visual-question-answering-c8c8093941bc>

Deep Learning for Visual Question Answering

<https://avisingh599.github.io/deeplearning/visual-qa/>

1. (Leonel, 2009) [↑](#footnote-ref-1)
2. (Fukui et al., n.d.) [↑](#footnote-ref-2)
3. טכניקה שהקבוצה המנצחת השתמשה בה [↑](#footnote-ref-3)
4. טכניקה שהקבוצה עם התוצאות הרעות ביותר השתמשה בה. [↑](#footnote-ref-4)
5. טכניקה שהקבוצה המנצחת השתמשה בה [↑](#footnote-ref-5)
6. טכניקה שהקבוצה המנצחת השתמשה בה [↑](#footnote-ref-6)