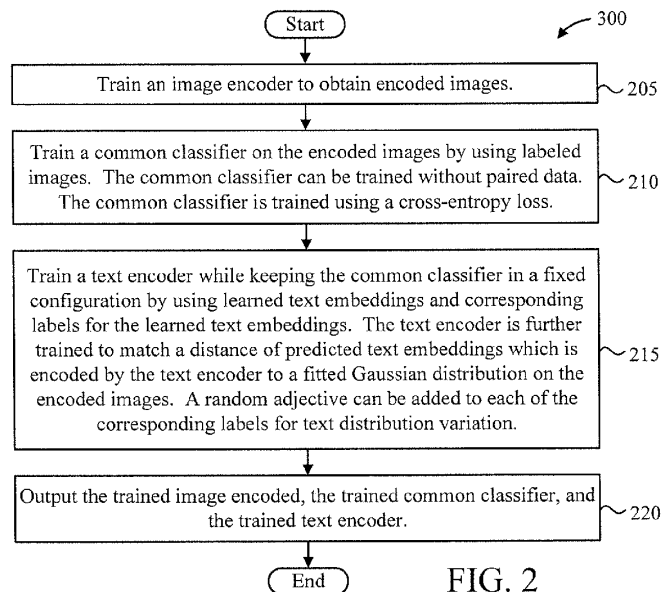




- (51) International Patent Classification:
G06K 9/62 (2022.01) *G06N 20/10* (2019.01)
- (21) International Application Number:
PCT/CN2021/128077
- (22) International Filing Date:
02 November 2021 (02.11.2021)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
17/109,550 02 December 2020 (02.12.2020) US
- (71) Applicant: **INTERNATIONAL BUSINESS MACHINES CORPORATION** [US/US]; New Orchard Road, Armonk, New York 10504 (US).
- (71) Applicant (for MG only): **IBM (CHINA) CO., LIMITED** [CN/CN]; 7F, Bldg. 10, Zhangjiang Innovation Park, 399 Keyuan Road, Zhangjiang High-Tech Campus, Pudong New Area, Shanghai 201203 (CN).

- (72) Inventors: **CHAUDHURY, Subhajt**; IBM Japan Ltd., 19-21 Nihonbashi-Hakozakicho, Chuo-Ku, Tokyo 103-8510 (JP). **KIMURA, Daiki**; IBM Japan Ltd., 19-21 Nihonbashi-Hakozakicho, Chuo-Ku, Tokyo 103-8510 (JP). **KURATA, Gakuto**; IBM Japan Ltd., 19-21 Nihonbashi-Hakozakicho, Chuo-Ku, Tokyo 103-8510 (JP). **TACHIBANA, Ryuki**; IBM Japan Ltd., 19-21 Nihonbashi-Hakozakicho, Chuo-Ku, Tokyo 103-8510 (JP).
- (74) Agent: **LIU, SHEN & ASSOCIATES**; 10th Floor, Building 1, 10 Caihefang Road, Haidian District, Beijing 100080 (CN).
- (81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW,

(54) Title: LEARNING UNPAIRED MULTIMODAL FEATURE MATCHING FOR SEMI-SUPERVISED LEARNING



(57) Abstract: A computer-implemented method is provided for learning multimodal feature matching. The method includes training an image encoder to obtain encoded images. The method further includes training a common classifier on the encoded images by using labeled images. The method also includes training a text encoder while keeping the common classifier in a fixed configuration by using learned text embeddings and corresponding labels for the learned text embeddings. The text encoder is further trained to match a distance of predicted text embeddings which is encoded by the text encoder to a fitted Gaussian distribution on the encoded images.

SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— with international search report (Art. 21(3))