# Çift Kollu Robotlarda Derin Deterministik Politika Gradyanı ile Hata Önleme
# Failure Prevention in Bimanual Robots using Deep Deterministic Policy Gradient

Asel Menekşe, Abdullah Cihan Ak, Sanem Sarıel
*Artificial Intelligence and Robotics Laboratory (AIR Lab)*
Istanbul Technical University, Istanbul, Turkey
{menekse16, akab, sariel}@itu.edu.tr

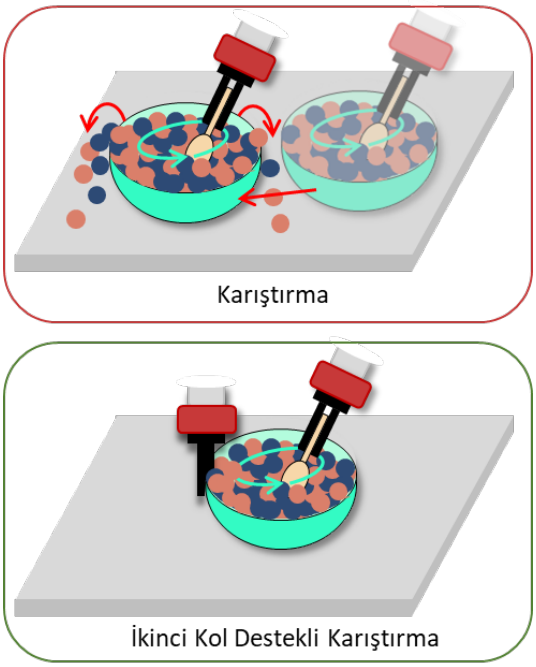
***Özetçe* —Robot nesne etkileşim davranışmalarının öğrenilmesinde karşılaşılan temel zorluklardan biri, hedefi emniyetli bir şekilde başarıya ulaştıran eniyi politikayı elde etmektir. Pekiştirmeli öğrenmede, hedefi ve emniyet kısıtlarını bir ödül fonksiyonu ile temsil etmek, genellikle emniyeti göz ardı eden çok temkinli olan bir politika öğrenmeye neden olabilir. Çünkü hedefe ulaşmak için yüksek ödüller sunan durumlar genellikle emniyetsiz durumlara daha yakındır. Bu bildiride, bu problemi aşmak için, genel görevi her bir robot kolu için atanmış iki alt göreve ayıran, emniyet kısıtlamalarını dikkate alan ve olası emniyetsiz durumları önleyen bir çift kol işbirliği yöntemi önerilmektedir. Bu yöntemde bir kol hedefi başarıyla gerçekleştirmeyi öğrenirken, diğer kolun bir ortam kısıtı tanımlanarak emniyeti sağlaması için çevreyle etkileşime geçmesi öğrenilir. Önerilen yöntem, benzetim ortamında insansı bir robotun bir kaseyi karıştırma görevini yerine getirmesi için uyarlanan 3 ayrı senaryo üzerinde sınanmış ve ikinci kol destekli karıştırma senaryosunun en iyi başarımı gösterdiği görülmüştür.**

***Anahtar Kelimeler*—*robot emniyeti, servis robotları, çift-kol işbirliği, pekiştirmeli öğrenme, çoklu-etmen sistemleri***

***Abstract*—One of the key challenges in learning robot object interaction behavior is to obtain an optimal policy that accomplishes the goal safely. In reinforcement learning, representing both the goal and safety constraints with a reward function may result in learning a policy that either omits safety or is too conservative; because states that yield high rewards for achieving the goal are often closer to unsafe zones. To overcome this problem, a dual-arm cooperation method is proposed, which decomposes the overall task into two subtasks assigned to each robot arm. This approach ensures that the goal is achieved while considering safety constraints and preventing potential unsafe situations. In this method, while one arm learns to achieve the goal successfully, the other arm learns to interact with the environment to ensure safety by defining an environmental constraint. The proposed method was tested in a simulation environment on 3 different scenarios adapted for a humanoid robot to perform the task of stirring a bowl, and it was seen that the second arm-assisted stirring scenario showed the best performance.**

***Keywords*—*robot safety, service robots, dual-arm cooperation, reinforcement learning, multi-agent systems***

**979-8-3503-8896-1/24/$31.00 ©2024 IEEE**




Şekil 1: Robotun karıştırma görevi esnasında oluşabilecek hataları önlemek için uygulayabileceği ikinci kol destekli karıştırma senaryosu.

## I. Giriş

Günümüzde robotlar, ortamla etkileşim yeteneklerinin artması ile birlikte günlük hayatımızda daha fazla yer almaya başlamışlardır [1], [2]. Fakat hala robotların ev veya mutfak ortamı gibi ortamlarda yemek pişirme ve ütü yapma gibi görevleri yürütmeleri tümüyle mümkün değildir. Özellikle robotların bu görevleri insanların bulunduğu ortamlarda yerine getirilebilmeleri için emniyetli şekilde yürütebilmeli, ortamda oluşabilecek emniyetsiz durumları tespit edebilmeli, tanımlayabilmeli [3] ve engellemeye yönelik yeteneklere sahip olmalıdır.

Robotlar güncel eniyileme ve öğrenme yöntemleri ile görevlerini başarıyla gerçekleştirebilmektedir. Ancak birçok başarı ölçütünü bir arada barındıran görevlerde başarım yeterli olamamaktadır. Robotun temel görevi öğrenmesi esnasında, eşzamanlı olarak emniyetsiz durumların oluşmasının önlenmesi, robotun ilgili görevi daha az yetkin şekilde öğrenmesine

sebep olabilmektedir. Bu nedenle, eylemlerin emniyetli şekilde yürütülmesi probleminin her hedef için ayrı yetenekler öğrenilmesi şeklinde temsil edilmesi, robot öğrenmesinin daha etkin olmasını sağlamaktadır [4].

Robotların bilişsel özelliklerinin geliştirilmesinde insanlardan ilham alınabilinir. Bir kaşık ile bir kasenin içindekileri karıştıran bir insan emniyetsiz durumları öngörme ve bunları önleme yeteneklerine sahiptir. Karıştırma sırasında bir insan genellikle birinci kolu ile karıştırma yaparken ikinci kolu ile sabitlemek için kaseyi tutar. Benzer şekilde, insansı bir robot için de ikinci kolunun, eylemlerin emniyetli yürütmesinde kullanışlı olduğu söylenebilir.

Bu çalışmada, robotların davranışlarını emniyetli olarak yürütecek şekilde öğrenmesi için ikinci robot kolunun kullanılması incelenmiştir. Yapılan çalışmada çift kollu insansı robotlarda birinci kolun ana görevi gerçekleştirecek eylemleri yürütmesi ve ikinci kolun emniyetsiz durumları önlemeye yönelik eylemleri yürütmesi ile emniyetli ortam sağlanmıştır. Ayrıca yetenek öğrenimi sırasında ikinci kolun aktif olarak emniyetsiz durumları önlemesinin, ana görevi daha başarılı gerçekleştiren yeteneklerin öğrenilmesini sağladığı gösterilmiştir. Deneylerde kullanılan örnek senaryo Şekil 1'de gösterilmiştir. Bu senaryoda robot karıştırma görevini gerçekleştirmektedir. Şekil 1'de üstteki resimde tek kolun öğrendiği karıştırma yeteneğinin yürütülmesi ile *kasenin kayması* ve *kasenin içindekilerin taşması* emniyetsiz durumlarının oluştuğu görülmektedir. Şekil 1'de alttaki resimde ikinci kolun desteği ile kasenin hareketi kısıtlanmıştır. Böylece ana görev için hem yetenek öğreniminde hem de eylem yürütmesinde emniyet sağlanmıştır.

CoppeliaSim benzetim ortamına yerleştirilen Baxter insansı robotu üzerinde yapılan deneylerde tespit edilen *kayma* ve *taşma* emniyetsiz durumlarının, etmen sayısı ikiye çıkarılarak önüne geçilmiş olup ikinci kolun hata önlemek için bir $c$ kısıtı öğrenmesi sağlanmasıyla birlikte kaseyi bir konuma götürme ve orada tutma yeteneği öğrenilmiştir. Bu yeteneğin ikinci kol ile kullanılmasının birinci kolun karıştırma yeteneğinin daha başarılı öğrenilmesini sağlarken olası *kayma* ve *taşma* hatalarının da önemli ölçüde azalttığı gösterilmiştir. Ayrıca her iki kol için belirtilen yeteneklerin beraber öğrenilmesinin de çoklu etmen yapılarına olan etkisi de bu çalışmada incelenmiştir.

## II. İLGİLİ ÇALIŞMALAR

Literatürde robotlarda derin pekiştirmeli öğrenme (Deep Reinforcement Learning, DRL) üzerine çeşitli makalaleler bulunmaktadır [5], [6]. Çift kollu robotlarda öğrenme üzerine yapılan önemli bir çalışmada [7], öğrenilmiş görev şemaları kullanılarak her iki robotik elin de kullanılmasını gerektiren görevleri yerine getirmek ve politikaları eğitmek için ham RGB görüntü gözlemlerini kullanırken, görevleri parametreli Markov karar süreçleri olarak temsil ederek seyrek ödüllere başa çıkma stratejisi benimsemekte ve öğrenilen becerilerin simülasyondan gerçek dünya senaryolarına etkili bir şekilde transfer edilmesine imkan tanınmaktadır. Bu, robotların şişe açma gibi karmaşık görevlerde verimli bir şekilde eğitilmesini olanak sağlamaktadır fakat oluşabilecek potansiyel hatalara karşı önlem alınması, hataların tahmini sorununa cevap verememektedir. Bu çalışmada ise, görevin öğrenmesi sağlanırken oluşabilecek hatalar bir ortam kısıtı ile giderilmiş olup daha uygulanabilir hale getirilirken model-bazlı bir pekiştirmeli öğrenme yöntemi ile robot eğitilmiştir.

Bir diğer çift kollu çalışmada [8] ise, insanlardan öğrenme gösterimleri (Learning from Demonstration, LfD) kullanarak robota insan benzeri bir şekilde wok tavasında Çin usulü karıştırma yapma becerisini optimal bir robot kontrol politikası ile öğretmeyi içerir. Şefler tarafından gösterilen gezingeleri taklit etmeyi hedefleyen bu politika, bir kolun hareketini temsil etmek için Dinamik Hareket Primitifleri'ni (Dynamic Motion Primitives, DMP) kullanır ve yeteneklerini artırmak için eğitilebilir otomatik bir zorlama terimi içerir. Bu, geleneksel kinematik tabanlı çift kollu koordinasyon yaklaşımlarından farklıdır.

Çift kollu robot kullanımı bağlamında parametreli eylemlere yönelik politika eğitimi yapan bir diğer çalışmada [7] ise, simülasyonda öğrenilen beceri dizilerini gerçek dünya görevlerine aktarmanın, görüntülerden seyrek ödül problemlerini verimli bir şekilde çözmeye olanak tanıyan ve gerçek robotların çift elleriyle işlem yürütme gibi karmaşık becerileri gerçekleştirmesi için uygulanabilir hale getirdiği sonuçlarına ulaşmış olmalarına rağmen güvenlikle alakalı herhangi bir kısıtı göz önünde bulundurmamışlardır. Bu çalışmada, ise bunun aksine güvenlik kısıtı göz önünde bulundurulup görevin başarıyla yürütülmesi sağlanmıştır.

Bunlara ek olarak, Liu ve diğerleri [8], derin takviyeli öğrenme modelini kullanarak bir nesneyi etkili bir şekilde kavrama olasılığını en üst düzeye çıkaracak uygun bir kavrama ve aynı zamanda optimum kavramayı bulma sorununu gidermek için robotik kavrama görevini bir Markov karar süreci (MDP) olarak formüle ederler ve Çift Derin Q-Öğrenme tabanlı bir mimari oluşturmuşlardır. Ayrıca çift kollu eylem yürütme bağlanımında bir kol cismi sabitlerken diğer kolun iş yürüttüğü bir çalışmada [9] ise, bir kol sabitlenecek noktayı görsel veri kullanarak öğrenip diğer kola iş kolaylığı sağlanmıştır. [9]'den farklı olarak derin pekiştirmeli öğrenme sonuçları verilmiştir.

Son zamanlarda yapılan bir çalışmada [10] ise, nesnelerle temas esnasında oluşabilecek hata senaryolarına karşı çekişmeli pekiştirme öğrenme yöntemi ile çoklu etmen yapıları için politikaların eğitimi ve bu hatalara karşı dirençli bir hata önleme politikasının geliştirilmesi üzerine gidilmiş ve etmenler rekabetçi etmenler ve bu etmenlere karşı başa çıkabilen bir baş eden etmenler tasarlanmıştır.

## III. ÇİFT KOLLU ROBOTLARDA HATA ÖNLEME

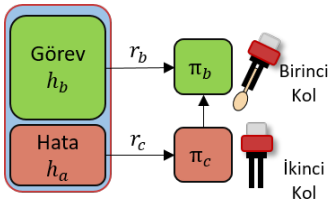
Bilişsel robotlar kendilerine verilen görevleri gerçekleştirmek için tasarlanmış veya öğrenilmiş motor yetenekleri kullanırlar. Sınırlandırılmış ortamlarda tasarlanan veya öğrenilen bu yetenekler gerçek robotlarla kullanıldığı zaman ortamın değişmesi ve daha önce deneyimlenmeyen durumların oluşması nedeniyle başarısız olabilmekte ve robotların kendileri ve ortamları için tehlikeli durumlara sebep verebilmektedirler. Bu nedenle robot öğrenmesi, verilen görevi gerçekleştirmeyi hedeflemekle birlikte ortamda oluşabilecek emniyetsiz durumları da önlemelidir.

Bu çalışmada, robot yetenek öğrenmesi problemi, Markov Karar Süreci(Markov Decision Process(MDP)) ile temsil edilmiştir. Bir robotun bulunduğu durum $s_t$'de robotun $a_t$ hareketini yapması ile robot $s_{t+1}$ durumuna geçmektedir. Bu geçiş robotun görevi göz önüne alınarak değerlendirilir ve robot ortamdan $r_t$ ödülünü gözlemler. Bu şekilde elde edilen ($s_t$, $a_t$, $r_t$, $s_{t+1}$) gözlemlerinin Pekiştirmeli öğrenme(Reinforcement

Learning) yöntemleri kullanılarak en iyilenmesi ile robot durumunu robot hareketine eşleyen politikalar öğrenilmektedir.

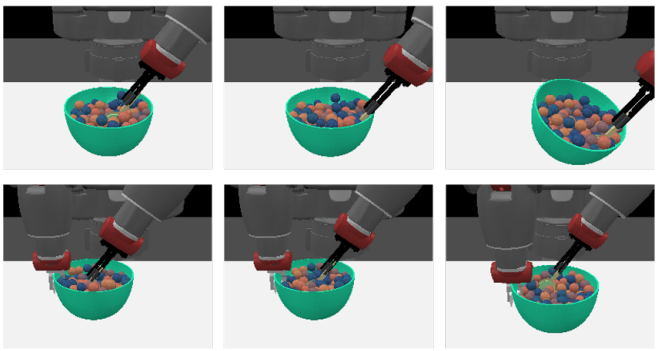
Robotlar kendilerine verilen görevleri gerçekleştirmek için öğrenilen politikaları $\pi_b$ baz motor yeteneği olarak tanımlanmaktadır. $\pi_b$ baz yeteneğinin başarılı öğrenilmesi ortamın karmaşıklığı ile bağlantılıdır. Robot öğrenmesi için belirlenen görev dışında robotun güvenli sınırlar içinde kalması gibi ek kriterler öğrenmenin etkinliğini düşürmektedir. Ortamda uygulanacak kısıtlamalar, etmen başına düşen hedef sayısını azaltmakla birlikte görev paylaşımına olanak sağlayıp etkin yetenek öğrenimini yerine getirmektedir.

Robotun görevi ile ilgili hedefi $h_b$ ve diğer güvenlik(hata önleme) hedefleri $h_a$ ile gösterilirse, robot ilgili hedef ödülü $r_b$ ve hata önleme ödülü $r_c$ ödülünü kullanarak $\pi_b$ baz yeteneğini RL ile öğrenmektedir. Ortama $h_a$ hedefini sağlayan bir kısıtlayıcı politika $\pi_c$ eklendiği zaman ortamın karmaşıklığı azalmaktadır ve $h_b$ hedefi için sadece $r_b$ ödülü ile $\pi_b$ baz yeteneği daha etkili öğrenilmektedir.




Şekil 2: Çift kollu güvenli robot öğrenme sistemi.

Bu çalışmada, robotlarda iki kollu emniyetli robot yürütmesi problemi iki kısımda gerçekleştirilmektedir. Önerilen yöntemde birinci kısım ikinci kol ile $h_a$ hedefini sağlayan $c$ kısıtını oluşturan $\pi_c$ yeteneğinin öğrenilmesidir. İkinci kısım öğrenilen $\pi_c$ yeteneğinin ikinci kol ile yürütülmesi problemin karmaşıklığı azaltılarak $\pi_b$ baz yeteneğinin öğrenilmesidir. $\pi_b$ ve $\pi_c$ yeteneklerinin öğrenilmesi ile bu iki yeteneğin ayrı kollarda aynı anda yürütülmesi ile güvenli robot yürütmesi sağlanmaktadır. Önerilen çift kollu güvenli robot öğrenmesi Şekil 2'de gösterilmiştir.



Şekil 3: Tek el ve çift el karıştırma senaryosu.

## IV. Deneyler

Bu bölümde, deney ortamı, deney tanımı ve eğitim süreci hakkında bilgi verilip eğitim esnasında oluşan hatalar gözlemlenmiş ve önerilen yönteme ilişkin sonuçlar karşılaştırılmalı sunulmuştur.

### A. Deney Ortamı ve Tanımı

Robot üzerinde yapılan deneyler, CoppeliaSim [11] benzetim ortamında Baxter insansı robotunun, masa üzerinde bulunan ve içinde farklı boyutlarda 180 top içeren kaseyi bir kaşık yardımıyla karıştırması senaryosu üzerinde gerçekleştirilmiştir.

Çalışmanın amacı, robotların belirli görevleri yerine getirirken sebep olabilecekleri emniyetsiz durumların nasıl giderilebileceğine yönelik bir çözüm üretimidir. Deneyde robotun karıştırma görevini yürütürken ortaya çıkabilecek kayma ve dökme hatalarının önüne geçebilmesi sınanmaktadır. Bunun için, bir önceki bölümde bahsedildiği üzere insansı robotun sahip olduğu iki kolun da kullanımının aktif olduğu bir sistem önerilmiştir. Bu sistemde ilk kol, baz hedef olan $h_b$ karıştırma görevini yerine getirirken, ikinci kol kasenin kayması ve içindekilerinin taşmasını engelleyen bir $c$ kısıtı ile $h_a$ hedefini sağlaması amaçlanmıştır. İkinci kolun oluşturduğu etkiyi gösterebilmek amacıyla 3 ayrı deney senaryosu oluşturulup her biri için ayrı eğitim yapılmıştır.

- **Tek Kol Karıştırma:** Bu senaryoda, kase ortamda serbest hareket edebilir durumdayken (sabitlenmemiş) sol elle karıştırma operasyonu öğretilir.
- **İkinci Kol Sabit Karıştırma:** Bu senaryoda, sağ kol için önceden eğitilmiş (pre-trained) bir model ile kaseyi sabitlerken, sol kol karıştırmayı öğrenir.
- **Çift Kol Öğrenme:** Bu senaryoda, sağ kol kaseyi hedef konumda tutmayı öğrenirken sol kol ise karıştırmayı öğrenir.

TABLO I: Durum Uzayı

| Durum | Vektör Gösterimi | Açıklama |
|---|---|---|
| $P_{kase}$ | $(x, y, z)$ | Kasenin konumu |
| $E_{sol-kol-eef}$ | $(x, y, z)$ | Sol kolun uç noktasının konumu (End Effector) |
| $E_{sag-kol-eef}$ | $(x, y, z)$ | Sağ kolun uç noktasının konumu (End Effector) |
| $D_{kase-sol-kol}$ | $(x, y, z)$ | Sol kolun uç noktasının kaseye göre konumu |
| $D_{kase-sag-kol}$ | $(x, y, z)$ | Sol kolun uç noktasının kaseye göre konumu |
| $P_{istenen}$ | $(x, y, z)$ | Kasenin istenen konumu |
| $D_{istenen}$ | $(x, y, z)$ | Kasenin, bulunması istenen noktaya göre konumu |
| $\theta$ | $(\theta)$ | Karıştırma fazı |

### B. Eğitim

Her iki kol da, bölüm sayısı 100 ve bölüm zaman adımı 100 olmak üzere toplam 10000 adım süresince eğitilir. Her bölüm başında, içerisinde farklı büyüklüklerde 180 topun olduğu kase rastgele bir konumda başlar.

Tek kol karıştırma senaryosunda, sol kol karıştırma görevini $\pi_b$ baz politikası (base policy) ile öğrenirken, sağ kol boş konumda bekler. İkinci kol sabit karıştırma senaryosunda ise, sol kol karıştırma görevini $\pi_b$ baz politikası (base policy) ile öğrenirken, sağ kol ise önceden eğittiğimiz $\pi_c$ kısıtlama politikası (constraint policy) ile $c$ kaseyi sabit tutar. Son olarak çift el karıştırma senaryosunda ise, aynı anda sol kol karıştırma görevini $\pi_b$ baz politikası (base policy) ile öğrenirken, sağ kol ise $\pi_c$ kısıtlama politikası(constraint policy) ile $c$ kısıtını öğrenip kaseyi sabit tutar. Üç politika da derin determistik

politika gradyanı yöntemi ile öğrenilmiş olup, Tablo I'de verilen durum uzayı $s_t$, eylem uzayı $[x, y, z]$ ile verilen $a_t$, olmak üzere kullanılan en iyi eylem-değer (optimal action-value) Bellman Denklemi (1) kullanılmıştır.

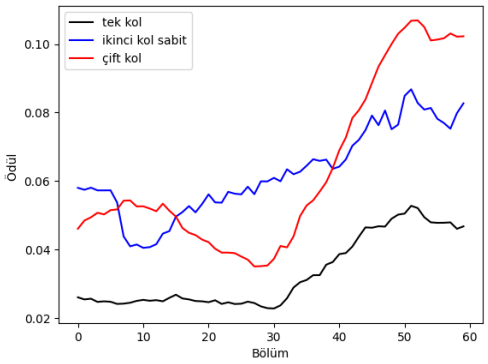
$$\pi^* (s_t, a_t) = E \left[ r (s_t, a_t) + \gamma \pi^* \left( s_{t+1}, a'_{t+1} \right) \right] \quad (1)$$

Sistemde karıştırma ödülü, dökme miktarı ve kayma miktarı olmak üzere 3 farklı ödül tanımlanmış olup $P'_{\text{toplar}}$, topların vektör biçiminde tutulan bir sonraki adımdaki konumlarını $[x_f, y_f, z_f]$ ile, $P_{\text{toplar}}$ ise şimdiki konumlarını $[x_c, y_c, z_c]$ ile, $B_{\text{taşma}}$ ise topların taşma miktarını ve n top sayısını temsil etmektedir. Denklem (2)'de verilen ödül fonksiyonları ile, sırasıyla karıştırma ödülü, kayma miktarı ve taşma miktarları her bir bölüm için hesaplanmıştır.

$$R_{genel} = \begin{cases} R_{karıştırma} = \sum_{i=0}^{n} \sqrt{(P'_{top,i} - P_{top,i})^2} \\ R_{kayma} = \sum \sqrt{(P'_{kase} - P_{kase})^2} \\ R_{taşma} = \sum B_{taşma} \end{cases} \quad (2)$$

### C. Deney Sonuçları

Her bir senaryo için bölüm zaman adımı 200 olmak üzere 60'şar bölümlük bir test yapılmıştır. Toplam 12000 zaman adımından oluşan bu testler, Şekil 4'de verilen ödül grafiğinde de görülebileceği üzere üç farklı deney senaryosu karşılaştırılmalı olarak verilmiştir. Siyah renkli grafikteki baz politika olan kasenin tek elle karıştırılması senaryosunun en düşük karıştırma ödülüne sahipken, mavi renkli grafikteki önceden eğitilmiş (pre-trained) sağ kol destekli karıştırma sonucunun daha iyi olduğu görülmektedir. Kırmızı renkli olan grafikteki iki kolun aynı anda eğitildiği senaryonun ise aldığı yüksek ödüle rağmen döktüğü top sayısının fazlalığı Tablo II'de görülebileceği üzere başarım oranını düşürmüştür.




Şekil 4: Ödül Grafiği

Tablo II'de 3 ayrı deney senaryosu için karıştırma ödülünün, ortalama kayma miktarının, maksimum kayma miktarının, ve taşma miktarlarının ortalama ve standart sapma değerleri verilmiştir. Tablodaki değerlerden de anlaşılabileceği üzere, en yüksek karıştırma ödülünü ikinci kol sabit senaryosu almıştır. Kayma miktarlarına bakılacak olursa, en başarılı senaryonun yine ikinci kol sabit senaryosunun olduğu görülecektir. Bunun sebebi öğrenilen $c$ ortam kısıtının sisteme eklendikten sonra diğer etmenin baz hedef olan $h_b$ karıştırma miktarını maksimize edebilmiş olmasıdır. Aynı anda öğrenme politikası olan çift kol senaryosunda ise bu karıştırma ödülünün tek kol karıştırma ödülünden fazlayken tek kol senaryosundaki taşma miktarının

TABLO II: Değerlendirme Sonuçları

| | Tek Kol | | İkinci Kol Sabit | | Çift Kol | |
|---|---|---|---|---|---|---|
| | ort. | std. | ort. | std. | ort. | std. |
| Karıştırma Ödülü | 0.03 | 0.01 | 0.14 | 0.04 | 0.10 | 0.07 |
| Ortalama Kayma | 0.11 | 0.03 | 0.01 | 0.00 | 0.21 | 0.05 |
| Maksimum Kayma | 0.18 | 0.03 | 0.02 | 0.01 | 0.25 | 0.05 |
| Taşma | 51.55 | 10.74 | 2.55 | 2.59 | 38.95 | 27.18 |

daha yüksek olduğu görülmüştür. Bunun sebebi, aynı anda iki etmenin ayrı hedefleri öğrenmesi ortam karmaşıklığını arttırarak bunu zor bir problem haline getirmesidir.

## V. Sonuç

Bu bildiride, derin pekiştirmeli öğrenme ile karmaşık ortamlarda deney yapmayı amaçlayan çift kollu robotlarda birinci kol ile istenen görev yerine getirilirken, olası güvenlik problemini önemli ölçüde azaltabilecek bir çözüm önerisiyle ikinci kolun kaseyi sabitlemesi ileri sürülmüş olup tek kolla karıştırma, ikinci kol sabit karıştırma, ve çift kol karıştırma senaryosu olmak üzere 3 ayrı deney senaryosu gerçekleştirilmiş olup deney sonuçları kısmında da verildiği üzere, robotun ikinci kol sabit karıştırma senaryosunda hataları en aza indirgediği ve karıştırma miktarı başarım oranının da en yüksek olduğu sonucuna ulaşılmıştır. Sim-to real uygulamalarıyla da gerçek dünyada uygulanabileceği düşünülmektedir.

## Bilgilendirme

Bu çalışma, Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından 119E-436 Numaralı proje ile desteklenmiştir. Projeye verdiği destekten ötürü TÜBİTAK'a teşekkürlerimizi sunarız.

## Kaynaklar

[1] A. S. Polydoros and L. Nalpantidis, "Survey of model-based reinforcement learning: Applications on robotics," *Journal of Intelligent & Robotic Systems*, vol. 86, no. 2, pp. 153–173, 2017.

[2] M. Ersen, E. Oztop, and S. Sariel, "Cognition-enabled robot manipulation in human environments: Requirements, recent work, and open problems," *IEEE Robotics Automation Magazine*, vol. 24, no. 3, pp. 108–122, 2017.

[3] A. Inceoglu, E. E. Aksoy, and S. Sariel, "Multimodal detection and classification of robot manipulation failures," *IEEE Robotics and Automation Letters*, vol. 9, no. 2, pp. 1396–1403, 2024.

[4] A. C. Ak, E. E. Aksoy, and S. Sariel, "Learning failure prevention skills for safe robot manipulation," *IEEE Robotics and Automation Letters*, vol. 8, no. 12, pp. 7994–8001, 2023.

[5] S. Joshi, S. Kumra, and F. Sahin, "Robotic grasping using deep reinforcement learning," 2020.

[6] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, and D. Wierstra, "Continuous control with deep reinforcement learning," 2019.

[7] R. Chitnis, S. Tulsiani, S. Gupta, and A. Gupta, "Efficient bimanual manipulation using learned task schemas," 2020.

[8] J. Liu, Y. Chen, Z. Dong, S. Wang, S. Calinon, M. Li, and F. Chen, "Robot cooking with stir-fry: Bimanual non-prehensile manipulation of semi-fluid objects," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, p. 5159–5166, Apr. 2022.

[9] J. Grannen, Y. Wu, B. Vu, and D. Sadigh, "Stabilize to act: Learning to coordinate for bimanual manipulation," 2023.

[10] M. C. Kutay, A. C. Ak, and S. Sariel, "Adversarial learning of failure prevention policies," in *The 31th Signal Processing and Communications Applications Conference (SIU)*, 2023, pp. 1–4.

[11] E. Rohmer, S. Singh, and M. Freese, "V-rep: A versatile and scalable robot simulation framework," 11 2013, pp. 1321–1326.